General Biology

Wikibooks.org

March 15, 2013

On the 28th of April 2012 the contents of the English as well as German Wikibooks and Wikipedia projects were licensed under Creative Commons Attribution-ShareAlike 3.0 Unported license. An URI to this license is given in the list of figures on page 175. If this document is a derived work from the contents of one of these projects and the content was still licensed by the project under this license at the time of derivation this document has to be licensed under the same, a similar or a compatible license, as stated in section 4b of the license. The list of contributors is included in chapter Contributors on page 169. The licenses GPL, LGPL and GFDL are included in chapter Licenses on page 179, since this book and/or parts of it may or may not be licensed under one or more of these licenses, and thus require inclusion of these licenses. The licenses of the figures are given in the list of figures on page 175. This PDF was generated by the LaTeX typesetting software. The LaTeX source code is included as an attachment (`source.7z.txt`) in this PDF file. To extract the source from the PDF file, we recommend the use of `http://www.pdflabs.com/tools/pdftk-the-pdf-toolkit/` utility or clicking the paper clip attachment symbol on the lower left of your PDF Viewer, selecting `Save Attachment`. After extracting it from the PDF file you have to rename it to `source.7z`. To uncompress the resulting archive we recommend the use of `http://www.7-zip.org/`. The LaTeX source itself was generated by a program written by Dirk Hünniger, which is freely available under an open source license from `http://de.wikibooks.org/wiki/Benutzer:Dirk_Huenniger/wb2pdf`. This distribution also contains a configured version of the `pdflatex` compiler with all necessary packages and fonts needed to compile the LaTeX source included in this PDF file.

Contents

1 Getting Started **3**

2 Biology - The Life Science **5**
2.1 Characteristics of life 5
2.2 Nature of science 6
2.3 Scientific method 6
2.4 Charles Darwin 9
2.5 After Darwin 9
2.6 Challenges to Darwin 10

3 The Nature of Molecules **11**
3.1 Matter 11
3.2 The atom 11
3.3 Mass and isotopes 12
3.4 Electrons 12
3.5 Chemical bonds 13
3.6 Chemical reactions 13
3.7 Water 13

4 The Chemical Building Blocks of Life **15**
4.1 Carbon 15
4.2 Carbohydrates 15
4.3 Stereoisomers 16
4.4 Lipids 16
4.5 Proteins 16
4.6 Hereditary (Genetic) information 18

5 Life: History and Origin **19**
5.1 Properties of life 19
5.2 Origin of life: 3 hypotheses 19
5.3 The early earth 20
5.4 Origin of cells 21
5.5 The RNA world? 21
5.6 The earliest cells 21
5.7 Major steps in evolution of life 22

6 Cells **23**

7 Cell structure **25**
7.1 What is a cell? 25
7.2 History of cell knowledge 29

7.3 Microscopes . . . 30
7.4 Cell size . . . 30

8 Structure of Eukaryotic cells 31
8.1 Structure of the nucleus . . . 31
8.2 Chromatin . . . 32
8.3 Endoplasmic reticulum . . . 32
8.4 The Golgi apparatus . . . 34
8.5 Ribosomes . . . 34
8.6 DNA-containing organelles . . . 34
8.7 Cytoskeleton . . . 35

9 Membranes 37
9.1 Biological membranes . . . 37
9.2 Phospholipid . . . 38
9.3 Fluid mosaic model . . . 38
9.4 Membrane proteins . . . 38
9.5 Receptor-mediated endocytosis . . . 40

10 Cell-cell interactions 41
10.1 Cell signaling . . . 41
10.2 Communicating junctions . . . 42

11 Energy and Metabolism 43
11.1 Energy . . . 43
11.2 Oxidation–Reduction . . . 43
11.3 NAD^+ . . . 44
11.4 Free energy . . . 44
11.5 Enzymes . . . 44
11.6 ATP . . . 46
11.7 Biochemical pathways . . . 46

12 Respiration: harvesting of energy 47
12.1 Energy . . . 47
12.2 Respiration . . . 47
12.3 Respiration of glucose . . . 47
12.4 Alternative anaerobic respiration . . . 47
12.5 Glycolysis overview . . . 48
12.6 Regeneration of NAD^+ . . . 48
12.7 Alcohol fermentation . . . 48
12.8 Lactate formation . . . 48
12.9 Krebs cycle: overview . . . 49
12.10 ATP production . . . 49
12.11 Evolution of aerobic respiration . . . 49

13 Photosynthesis 51
13.1 Light Reactions . . . 51
13.2 “Dark” reactions . . . 53
13.3 Prokaryote cell division . . . 53

13.4 Bacterial DNA replication . . . 54
13.5 Chromosome number . . . 54
13.6 Eukaryotic chromosomes . . . 54
13.7 Chromosome organization . . . 55
13.8 Human karyotype stained by chromosome painting . . . 55
13.9 Chromosomes . . . 55
13.10 Human chromosomes . . . 55
13.11 Mitotic cell cycle . . . 55
13.12 Replicated human chromosomes . . . 56
13.13 Mitosis . . . 56
13.14 Plant mitosis . . . 56
13.15 Controlling the cell cycle . . . 56
13.16 Cancer . . . 57
13.17 Mutations and cancer . . . 57

14 Sexual reproduction 59
14.1 Sexual . . . 59
14.2 Sexual life cycle . . . 59
14.3 Meiosis . . . 59
14.4 Prophase I: synapsis . . . 60
14.5 Crossing over . . . 60
14.6 Microtubules and anaphase I . . . 60
14.7 Meiosis II . . . 60
14.8 Evolution of sex . . . 60
14.9 Consequences of sex . . . 61

15 Genetics 63

16 Gregor Mendel and biological inheritance 65
16.1 Mendel . . . 65
16.2 Mendel's experiments . . . 66
16.3 Mendel's seven pairs of traits . . . 66
16.4 Locus . . . 66
16.5 Modern Y chromosome . . . 67
16.6 Chromosome phenomena . . . 67
16.7 X-chromosome inactivation . . . 67
16.8 Barr body . . . 67
16.9 Human genetic disorders . . . 67

17 DNA: The Genetic Material 69
17.1 DNA . . . 69
17.2 Historical perspective . . . 69
17.3 Hershey-Chase Experiment . . . 69
17.4 DNA/RNA components . . . 70
17.5 Chemical structure of DNA . . . 70
17.6 3D structure of DNA . . . 70
17.7 Franklin . . . 70
17.8 DNA replication . . . 71

17.9 DNA replication . . . 71
17.10 DNA polymerases . . . 71
17.11 DNA replication complex . . . 72
17.12 DNA replication . . . 72
17.13 DNA replication fork . . . 72
17.14 Replication units . . . 72
17.15 Replicon . . . 72
17.16 What is gene? . . . 72

18 Gene expression **75**
18.1 "Central Dogma" . . . 75
18.2 The Genetic Code . . . 75
18.3 Transcription . . . 75
18.4 Transcription bubble . . . 76
18.5 Eukaryote mRNA . . . 76
18.6 Translation . . . 76
18.7 Translation in bacteria . . . 76
18.8 Aminoacyl tRNA synthase . . . 77
18.9 Ribosome structure . . . 77
18.10 Large ribosome subunit . . . 77
18.11 Translation . . . 77
18.12 Initiation complex . . . 77
18.13 Elongation, translocation . . . 77
18.14 Introns/exons . . . 77

19 Gene regulation **79**
19.1 Transcriptional control . . . 79
19.2 DNA grooves . . . 79
19.3 Regulatory proteins . . . 80
19.4 Lac operon of E. coli . . . 80
19.5 Alternative splicing . . . 80

20 Mutation **81**
20.1 Point Mutations . . . 81
20.2 Substitution . . . 81
20.3 Larger mutations . . . 82
20.4 Chromosomal mutations . . . 82
20.5 Causes of mutations . . . 82
20.6 Effects of mutations . . . 82
20.7 Further reading . . . 83
20.8 Original notes . . . 83
20.9 Point mutation . . . 84
20.10 Acquisition of genetic variability . . . 84
20.11 Eukaryote genome . . . 84
20.12 Barbara McClintock . . . 85

21 Recombinant DNA technology **87**
21.1 Recombinant DNA technology . . . 87

21.2 Restriction endonucleases . . . 87
21.3 Restriction endonucleases . . . 88
21.4 Uses of cloned gene . . . 88
21.5 Other molecular procedures . . . 88
21.6 RFLP(restriction fragment length polymorphism) analysis . . . 89
21.7 Sanger DNA sequencing . . . 89
21.8 Automated sequencing . . . 89
21.9 Genome projects . . . 89
21.10 Biochips . . . 90
21.11 DNA chip controversies . . . 90
21.12 Gene patenting . . . 90
21.13 Stem cells . . . 91

22 Classification of Living Things 93
22.1 Introduction . . . 97
22.2 Viral Replication . . . 98
22.3 Viral Genome . . . 99
22.4 Viruses Practice Questions . . . 99
22.5 Archaea . . . 100
22.6 Prokaryote evolution . . . 100
22.7 Domains of life: characteristics . . . 101
22.8 Introduction . . . 101
22.9 Classification of Protists . . . 101
22.10 Protozoa . . . 101
22.11 Algae . . . 102
22.12 Slime molds & Water molds . . . 103
22.13 Protists Practice Questions . . . 104

23 Multicellular Photosynthetic Autotrophs 107
23.1 Plants . . . 107
23.2 Plant phyla . . . 107
23.3 Plant evolution . . . 107
23.4 Plant phylogeny . . . 108
23.5 Plant life cycles . . . 108
23.6 Moss life cycle . . . 108
23.7 Vascular plants . . . 108
23.8 Vascular plant life cycles . . . 109
23.9 Pterophyta (ferns) . . . 109
23.10 Non-seed plants, continued . . . 109
23.11 Seed plants . . . 109
23.12 Sporophyte/gametophyte . . . 110
23.13 Megasporangium (nucellus) . . . 110
23.14 Pollen . . . 110
23.15 Gymnosperms . . . 110
23.16 Pine life cycle . . . 110
23.17 Other Coniferophyta . . . 110
23.18 Other gymnosperms . . . 111
23.19 Angiosperms . . . 111

23.20 Earliest angiosperm . . . 111
23.21 Angiosperm flower . . . 111
23.22 Angiosperm life cycle . . . 111
23.23 Introduction . . . 111
23.24 Nutrition . . . 112
23.25 Fungal Reproduction . . . 112
23.26 Types of Fungi . . . 113
23.27 Key Terms . . . 114
23.28 Introduction . . . 114
23.29 Characteristics of an Animal . . . 114
23.30 Introduction to animal phyla . . . 116
23.31 Phylum Porifera . . . 117
23.32 Phylum Cnidaria . . . 118
23.33 Phylum Platyhelminthes . . . 120
23.34 Phylum Rotifera . . . 121
23.35 Phylum Nematoda . . . 121
23.36 Phylum Annelida . . . 122
23.37 Phylum Arthropoda . . . 123
23.38 Phylum Mollusca . . . 124
23.39 Phylum Echinodermata . . . 124
23.40 Phylum Chordata . . . 125

24 Chordates 127
24.1 Characteristics . . . 127
24.2 Subphylum Urochordata . . . 127
24.3 Subphylum Cephalochordata . . . 128
24.4 Subphylum Vertebrata . . . 128

25 Tissues and Systems 135

26 Epithelial tissue 137

27 Connective tissue 139

28 Muscle tissue 143

29 Vertebrate digestive system 147

30 Circulatory system 151

31 Respiratory system 155
31.1 Neuron structure . . . 156
31.2 Central nervous system . . . 157
31.3 Peripheral nervous system . . . 157

32 Sensory systems 159
32.1 Taste and smell (chemoreception) . . . 159
32.2 Response to gravity and movement . . . 159
32.3 Vision . . . 160

32.4 Homeostasis 160
32.5 Osmotic environments and regulations 161

33 Additional material **165**
33.1 External Links 166

34 Glossary **167**
34.1 Users 168

35 Contributors **169**

List of Figures **175**

36 Licenses **179**
36.1 GNU GENERAL PUBLIC LICENSE 179
36.2 GNU Free Documentation License 180
36.3 GNU Lesser General Public License 181

1 Getting Started

2 Biology - The Life Science

The word ***biology*** means, "the science of life", from the Greek **bios,** *life*, and **logos,** *word* or *knowledge.* Therefore, Biology is the science of Living Things. That is why Biology is sometimes known as Life Science.

The science has been divided into many subdisciplines, such as botany[1], bacteriology, anatomy[2], zoology, histology, mycology, embryology, parasitology, genetics[3], molecular biology[4], systematics, immunology, microbiology[5], physiology, cell biology[6], cytology, ecology[7], and virology. Other branches of science include or are comprised in part of biology studies, including paleontology[8], taxonomy, evolution, phycology, helimentology, protozoology, entomology, biochemistry, biophysics, biomathematics, bio engineering, bio climatology and anthropology.

2.1 Characteristics of life

Not all scientists agree on the definition of just what makes up life. Various characteristics describe most living things. However, with most of the characteristics listed below we can think of one or more examples that would seem to break the rule, with something nonliving being classified as living or something living classified as nonliving. Therefore we are careful not to be too dogmatic in our attempt to explain which things are living or nonliving.

- Living things are composed of **matter structured in an orderly way** where simple molecules are ordered together into much larger macromolecules.

An easy way to remember this is GRIMNERD C All organisms; - **G**row, **R**espire, **I**nteract, **M**ove, Need **N**utrients, **E**xcrete (Waste), **R**eproduce,**D**eath, **C**ells (Made of)

- Living things are **sensitive,** meaning they are able to **respond to stimuli.**
- Living things are able to **grow**, **develop**, and **reproduce**.
- Living things are able to **adapt** over time by the process of **natural selection**.
- All known living things use the **hereditary molecule, DNA**[9].

1 http://en.wikibooks.org/wiki/botany
2 http://en.wikibooks.org/wiki/anatomy
3 http://en.wikibooks.org/wiki/genetics
4 http://en.wikibooks.org/wiki/Molecular%20Biology
5 http://en.wikibooks.org/wiki/microbiology
6 http://en.wikibooks.org/wiki/Cell%20Biology
7 http://en.wikibooks.org/wiki/ecology
8 http://en.wikibooks.org/wiki/paleontology
9 http://en.wikipedia.org/wiki/DNA

- Internal functions are coordinated and **regulated** so that the internal environment of a living thing is relatively constant, referred to as **homeostasis[10].**

Living things are organized in the microscopic level from atoms up to cells[11]. Atoms are arranged into molecules, then into macromolecules[12], which make up organelles[13], which work together to form cells. Beyond this, cells are organized in higher levels to form entire multicellular organisms. Cells together form tissues[14], which make up organs, which are part of organ systems, which work together to form an entire organism. Of course, beyond this, organisms form populations which make up parts of an ecosystem. All of the Earth's ecosystems together form the diverse environment that is the earth.

Example:-

sub atoms, atoms, molecules, cells, tissues, organs, organ systems, organisms, population, community, eco systems

2.2 Nature of science

Science is a **methodology** for **learning about the world**. It involves the **application of knowledge**.

The scientific method deals with **systematic investigation**, **reproducible results**, the formation and testing of **hypotheses**, and **reasoning.**

Reasoning can be broken down into two categories, **induction** (specific data is used to develop a generalized observation or conclusion) and **deduction** (general information leads to specific conclusion). Most reasoning in science is done through induction.

Science as we now know it arose as a discipline in the 17th century.

2.3 Scientific method

The scientific method is not a step by step, linear process. It is an intuitive process, a methodology for learning about the world through the application of knowledge. Scientists must be able to have an "imaginative preconception" of what the truth is. Scientists will often observe and then hypothesize the reason why a phenomenon occurred. They use all of their knowledge and a bit of imagination, all in an attempt to uncover something that might be true. A typical scientific investigation might go like so:

> You *observe* that a room appears dark, and you ponder *why* the room is dark. In an attempt to find explanations to this curiosity, your mind unravels several different *hypotheses*. One hypothesis might state that the lights are turned off. Another hunch might be that the room's lightbulb has burnt out. Worst yet, you could be going blind. To discover the truth,

10 http://en.wikipedia.org/wiki/homeostasis
11 http://en.wikipedia.org/wiki/cell
12 http://en.wikipedia.org/wiki/macromolecule
13 http://en.wikipedia.org/wiki/organelle
14 http://en.wikibooks.org/wiki/General%20Biology%2FTissues

you *experiment*. You feel your way around the room and find a light switch and turn it on. No light. You *repeat* the experiment, flicking the switch back and forth. Still nothing. That means your initial hypothesis, the room is dark because the lights are off, has been *rejected*. You devise more experiments to test your hypotheses, utilizing a flashlight to prove that you are indeed not blind. In order to *accept* your last remaining hypothesis as the truth, you could *predict* that changing the light bulb will fix the problem. If all your predictions succeed, the original hypothesis is valid and is accepted. In some cases, however, your predictions will not occur, in which you'll have to start over. Perhaps the power is off.

Figure 1 How Science is Done
A diagram that illustrates scientific investigation

Scientists first make observations that raise a particular question. In order to explain the observed phenomenon, they develop a number of possible explanations, or hypotheses. This is the inductive part of science, observing and constructing plausible arguments for why

an event occurred. Experiments are then used to eliminate one of more of the possible hypotheses until one hypothesis remains. Using deduction, scientists use the principles of their hypothesis to make predictions, and then test to make sure that their predictions are confirmed. After many trials (repeatability) and all predictions have been confirmed, the hypothesis then may become a theory.

Quick Definitions

Observation - Quantitative and qualitative measurements of the world.

Inference - Deriving new knowledge based upon old knowledge.

Hypotheses - A suggested explanation.

Rejected Hypothesis - An explanation that has been ruled out through experimentation.

Accepted Hypothesis - An explanation that has not been ruled out through excessive experimentation and makes verifiable predictions that are true.

Experiment - A test that is used to rule out a hypothesis or validate something already known.

Scientific Method - The process of scientific investigation.

Theory - A widely accepted hypothesis that stands the test of time. Often tested, and usually never rejected.

The scientific method is based primarily on the testing of hypotheses by experimentation. This involves a **control**, or subject that does not undergo the process in question. A scientist will also seek to limit variables to one or another very small number, single or minimum number of variables. The procedure is to form a hypothesis or prediction about what you believe or expect to see and then do everything you can to violate that, or falsify the hypotheses. Although this may seem unintuitive, the process serves to establish more firmly what is and what is not true.

A founding principle in science is a lack of absolute truth: the accepted explanation is the most likely and is the basis for further hypotheses as well as for falsification. All knowledge has its relative uncertainty.

Theories are hypotheses which have withstood repeated attempts at falsification. Common theories include evolution by natural selection and the idea that all organisms consist of cells. The scientific community asserts that much more evidence supports these two ideas than contradicts them.

2.4 Charles Darwin

Figure 2

Charles Darwin is most remembered today for his contribution of the theory of **evolution through natural selection**.

The seeds of this theory were planted in Darwin's mind through observations made on a five-year voyage through the New World on a ship called the Beagle. There, he studied fossils and the geological record, geographic distribution of organisms, the uniqueness and relatedness of island life forms, and the affinity of island forms to mainland forms.

Upon his return to England, Darwin pondered over his observations and concluded that evolution must occur through natural selection. He declined, however, to publish his work because of its controversial nature. However, when another scientist, Wallace, reached similar conclusions, Darwin was convinced to publish his observations in 1859. His hypothesis revolutionized biology and has yet to be falsified by empirical data collected by mainstream scientists.

2.5 After Darwin

Since Darwin's day, scientists have amassed a more complete **fossil record**, including **microorganisms** and **chemical fossils**. These fossils have supported and added subtleties to Darwin's theories. However, the age of the Earth is now held to be much older than Darwin thought. Researchers have also uncovered some of the preliminary mysteries of the mechanism of heredity as carried out through **genetics** and **DNA**, areas unknown to Darwin. Another growing area is **comparative anatomy** including homology and analogy.

Today we can see a bit of evolutionary history in the **development** of embryos, as certain (although not all) aspects of development recapitulate evolutionary history.

The molecular biology[15] study of slowly mutating genes reveal considerable evolutionary history consistent with fossil and anatomical record.

2.6 Challenges to Darwin

Figure 3

Darwin and his theories have been challenged many times in the last 150 years. The challenges have been primarily religious based on a perceived conflict with the preconceived notion of creationism. Many of those who challenge Darwin have been adherents to the young earth hypothesis that says that the Earth is only some 6000 years old and that all species were individually created by a god. Some of the proponents of these theories have suggested that chemical and physical laws that exist today were different or nonexistent in earlier ages. However, for the most part, these theories are either not scientifically testable and fall outside the area of attention of the field of biology, or have been disproved by one or more fields of science.

This text is based on notes very generously donated by Dr. Paul Doerder, Ph.D., of Cleveland State University.

15 http://en.wikibooks.org/wiki/molecular%20biology

3 The Nature of Molecules

3.1 Matter

Matter is defined as anything that has mass[1] (an amount of matter in an object) and occupies space[2] (which is measured as volume[3]).

- Particles, from smallest to largest
  1. Subatomic particles
     - Electrons[4]
     - Protons[5]
     - Neutrons[6]
  2. Atoms
  3. Molecules
  4. Macromolecules
- Origin of matter
  1. Big Bang[7], about 13.7 billion years ago
  2. Hydrogen[8], helium[9]
  3. Heavier elements formed in suns, super nova
     - Earth's matter predates formation of sun, 4.5 billion years ago
- All matter consists of atoms, which are composed of : electrons, protons, neutrons

3.2 The atom

- Example: Hydrogen
  - The simplest element
  - One proton (+)
  - One electron in orbit (-)
- Built by adding one proton (and one electron) at a time
- Number of protons determines atomic number and number of electrons
- Neutrons
  - Neutral charge

---

1 http://en.wikipedia.org/wiki/mass
2 http://en.wikipedia.org/wiki/space%23Physics
3 http://en.wikipedia.org/wiki/volume
4 http://en.wikipedia.org/wiki/Electrons
5 http://en.wikipedia.org/wiki/Protons
6 http://en.wikipedia.org/wiki/Neutrons
7 http://en.wikipedia.org/wiki/Big%20Bang
8 http://en.wikipedia.org/wiki/Hydrogen
9 http://en.wikipedia.org/wiki/Helium

    - Contribute mass
    - May decay
- Oxygen[10]
    - 8 protons (mass)
    - 8 electrons
    - 8 neutrons (mass)

3.3 Mass and isotopes

- Atomic mass
    - Sum of masses of protons and neutrons
    - Measured in daltons or AMU (Atomic Mass Unit)
    - An AMU is 1/12 the mass of Carbon-12
    - proton ~1 AMU or dalton
    - 6.024 x 10^{23} daltons/gram
- Atoms with same atomic number belong to same element
- Isotopes
    - Same atomic number but different atomic mass
    - Some are radioactive
    - Uses of isotopes
        - Radioactive: 3H, 14C, 32P, 35S
            - Tracers in biochemical reactions
            - Detection of molecules in recombinant DNA technology (genetic engineering)
            - Half-life: dating of rocks, fossils
        - Non-radioactive (N, C, O)
            - Diet of organisms (including fossils)
            - Biochemical tracers

3.4 Electrons

- Negative charge
- Held in orbit about nucleus by attraction to positively charged nucleus
- Atom may gain or lose electron, altering charge
    - Cation: loses electron, positive charge
        - Na^+
    - Anion: gains electron, negative charge
        - Cl^-
- Determine chemical properties of atoms
    - Number
    - Energy level

10 http://en.wikipedia.org/wiki/Oxygen

3.5 Chemical bonds

- Form molecules
- Enzymes: make, break, rearrange chemical bonds in living systems
- Ionic
- Covalent
  - Sharing of one or more pairs of electrons
    - Called single, double, or triple
  - No net charge (as in ionic bonds)
  - No free electrons
  - Give rise to discrete molecules
  - Hydrogen

3.6 Chemical reactions

- Formation and breaking of chemical bonds
- Shifting arrangement of atoms
- Reactants -> products
- Reactions are influenced by:
  - Temperature
  - Concentration of reactants, products
  - Presence of catalysts (enzymes)
- Oxidation:reduction

3.7 Water

- Essential for life
- ~75% earth's surface is water
- Life evolved in water
- Solvent for many types of solutes
- High specific heat
- High polarity
  - Creates a slightly negative Oxygen and a Slightly positive hydrogen
  - allows formation of Hydrogen Bonds

3.7.1 Hydrogen bonding

- A type of polar interaction
- Critical for:
  - Protein structure
  - Enzymatic reactions
  - Movement of water in plant stems
- Weak and transient
- Powerful cumulative effect

  - Solubility of many compounds
  - Cohesion (capillary action)
  - Lower density of ice
- Formed between molecules other than water
  - Protein structure
  - DNA[11], RNA[12] structure

Water organizes nonpolar molecules

- Nonpolar molecules: no polarity (+/-) charges
- Hydrophobic: exclude water because they don't form hydrogen bonds with it
- Consequences:
  - Membranes
  - Protein structure
- Hydrophilic: polar substances associate with water

Ionization of water: H_2O -> H^+ + OH^-

- Forms a Hydrogen ion (H^+), hydroxide ion (OH^-)
- Due to spontaneous breakage of covalent bond
- At 25°C, 1 liter of water contains 10-7 moles of H^+ ions: 10-7 moles/liter

pH

- A convenient way of indicating H^+ concentration
- pH[13] = -log[H^+]
- For water, pH = -log[10-7] = 7
- Since for each H^+ in pure water, there is one OH^-, pH of 7 indicates neutrality
- Logarithmic scale

Buffer

- Reservoir for H^+
- Maintains relatively constant pH over buffering range

This text is based on notes very generously donated by Dr. Paul Doerder, Ph.D., of the Cleveland State University.

11 http://en.wikipedia.org/wiki/DNA
12 http://en.wikipedia.org/wiki/RNA
13 http://en.wikipedia.org/wiki/pH

4 The Chemical Building Blocks of Life

Building blocks of life

- Carbon based: organic molecules
- Carbohydrates: CHO
- Lipids: CHO, water insoluble
- Proteins: CHONS, structure/function in cells
- Nucleic acids: CHONP, hereditary (genetic) information

4.1 Carbon

- Can make 4 covalent bonds[1]
  - Chains
    - Straight
    - Branched
    - Ring
  - Hydrocarbons[2] (C, H): store energy
  - Functional groups
    - Attach to carbon
    - Alter chemical properties
    - Form macromolecules
    - Sapoteton

4.2 Carbohydrates

- Principally CHO (rare N, S and P)
  - 1C:2H:1O ratio
  - Energy rich (many C-H bonds)
- Monosaccharides (principal: glucose[3])
  - Simple sugars
  - Principle formula: $C_6H_{12}O_6$
  - Form rings in water solution
- Disaccharides (sucrose, lactose)
- Polysaccharides (starch, glycogen, cellulose, chitin)

1 http://en.wikipedia.org/wiki/covalent%20bonds
2 http://en.wikipedia.org/wiki/Hydrocarbons
3 http://en.wikipedia.org/wiki/glucose

4.3 Stereoisomers

- Bond angles of carbon point to corners of a tetrahedron
- When 4 different groups are attached to a carbon, it is asymmetric, leading to various types of isomerism
  - Stereoisomers: (D, L)
- Same chemical properties
- Different biological properties
- D sugars, L amino acids

4.4 Lipids

- C-H bonds (nonpolar) instead of C-OH bonds as in carbohydrates
  - High energy
  - Hydrophobic (insoluble in water)
- Categories
  - Fats: glycerol and three fatty acids
  - Phospholipids: primary component of membranes
  - Prostaglandins: chemical messengers (hormones)
  - Steroids: membrane component; hormones
  - Terpenes: pigments; structure

4.4.1 Fatty acids

- Hydrocarbon chain
  - Even number of C, 14->20
  - Terminates in carboxyl group
- Saturated: contain maximum number of hydrogens (all single bonds); maximum energy
- Unsaturated: one or more double bonds
  - Usually higher melting point
  - Many common oils are polyunsaturated

4.5 Proteins

- Polymer of amino acids
  - 21 different amino acids found in proteins
  - Sequence of amino acids determined by gene
- Amino acid sequence determines shape of molecule
  - Linked by peptide bond (covalent)
- Functions
  - regulate chemical reactions and cell processes [enzymes]
  - form bone and muscle; various other tissues
  - facilitate transport across cell membrane [carrier proteins]

  - fight disease [antibodies]
- Motifs: folding patterns of secondary structure
- Domains: structural, functional part of protein often independent of another part; often encoded by different exons
- Shape determines protein's function

4.5.1 Amino acids

- 21 commonly found in proteins
  - 21st is selenocysteine, not mentioned in text
- Common structure
  - Amino group: NH_2
  - Carboxyl group: COOH
  - R group- 4 different kinds of R groups
    - acidic
    - basic
    - hydrophilic (polar)
    - hydrophobic (nonpolar)
- Confer individual properties on amino acids
- List of amino acids[4]

4.5.2 Structure

- Primary structure: the amino acid sequence
  - Determines higher orders of structure
  - Critical for structure and function of protein
- Secondary: stabilized by intramolecular hydrogen bonding
  - helix
  - sheet
- Tertiary: folding, stabilized by ionic bonds (between R groups), hydrogen bonding, van der Waal's forces, hydrophobic interactions
- Quaternary: _2 polypeptides

4.5.3 Function

- Requires proper folding, cofactors, pH, temperature, etc.
- Proteins are often modified after synthesis
  - Chemical modification
  - Addition of heme groups (hemoglobin, cytochrome)
- Denatured proteins can not function properly
- Proteins are degraded by proteosome as part of constant turnover of cell components

4 http://en.wikipedia.org/wiki/Amino%20acid%23List%20of%20amino%20acids

4.6 Hereditary (Genetic) information

- Nucleic acids
  - DNA: deoxyribonucleic acid
- Hereditary information of all cells
- Hereditary information for many viruses
  - RNA: ribonucleic acid
- Hereditary information of certain viruses (HIV[5])
- Intermediate in gene expression
- Composed of nucleotides
  - Ribonucleotides
  - Deoxyribonucleotides

4.6.1 RNA DNA origin

- Which came first?
- Paradox: DNA encodes protein necessary for its own replication
- Discovery of catalytic RNA by Cech and Altman suggested that RNA might have been first self-replicating molecule
- DNA evolved as more stable type of storage molecule

This text is based on notes very generously donated by Dr. Paul Doerder, Ph.D., of the Cleveland State University.

Proteins: Their building block is amino acids. The bond connecting 2 of the amino acids together are called peptide bonds. One of these bonds makes a monopeptide, two a dipeptide, and any more than that makes a polypeptide.

5 http://en.wikipedia.org/wiki/HIV

5 Life: History and Origin

5.1 Properties of life

1. Organization: Being structurally composed of one or more cells, which are the basic units of life.
   - prokaryote: no nucleus
   - eukaryote: membrane bound nucleus.

2. Sensitivity: respond to stimuli.

3. Energy Processing

4. Growth and Development

5. Reproduction

   - hereditary mechanisms to make more of self; DNA based.
6. Regulation, including homeostasis.

7. Evolution.

5.2 Origin of life: 3 hypotheses

- Extraterrestrial origin (panspermia): meteor, comet borne from elsewhere in universe
  - evidence of amino acids and other organic material in space (but often both D & L forms)
  - questionable bacterial fossils in Martian rock

-However, this would imply that some other origin of life was likely because it would have had to happen elsewhere before it could be transported here, and the only difference would be that life did not originate on Earth.

- Spontaneous origin on earth: primitive self-replicating macromolecules acted upon by natural selection ((macro)Evolution is one example of this)

-This is often attacked for the seeming impossibility for life to have been produced by a chemical reaction triggered by lightning and the ability of any produced DNA to actually be in a sequence that could produce a working model of life if replicated. It is also attacked for religious reasons, as it bypasses things like the idea of a supreme being directly creating humans. It also seems unlikely to some that such huge changes are possible in evolution

without evidence of an "in-between stage" that is credible. Many of the stages of man are disputed due to their somewhat shakey grounds. For example, bones from other animals have been taken accidentally in some cases to be part of a humanoid, and complete skeletons have been sketched out from a limited number of bones.

- Special creation: religious explanations (Intelligent Design is one popular example of this.) These explanations contend that life was created by God (or perhaps some other Intelligent Designer).
  - Proponents of Intelligent design suggest that the vast complexity of life could only have been intentionally designed while other creationists cite biblical support.

-This is often attacked for many of the same reasons that religion is attacked, and is often regarded as superstitious and/or unscientific.

- It is debated as to whether schools should teach one hypothesis or the other when talking about the origin of life. However, since they are all currently known major hypotheses (and sometimes hypotheses proven wrong are shown for educational purposes), this wikibook includes what it can without discriminating unfairly against one hypothesis or the other.

5.3 The early earth

It is believed that the Earth was formed about 4.5 billion years ago.

- Heavy bombardment by rubble ceased about 3.8 billion years ago.
- Reducing atmosphere: much free H
  - also H_2O, NH_3, CH_4
  - little, if any, free O_2
  - with numerous H electrons, require little energy to form organic compounds with C
- Warm oceans, estimated at 49-88°C
- Lack of O_2 and consequent ozone (O_3) meant considerable UV energy

Chemical reactions on early earth

- UV and other energy sources would promote chemical reactions and formation of organic molecules
- Testable hypothesis: Miller-Urey experiment
  - simulated early atmospheric conditions
  - found amino acids, sugars, etc., building blocks of life
  - won Nobel prize for work
  - experiment showed prebiotic synthesis of biological molecules was possible

Issues

- Miller later conceded that the conditions in his experiments were not representative of what is currently thought to be those of early earth
- He also conceded that science has no answer for how amino acids could self-organize into replicating molecules and cells
- In the 50 years since Miller-Urey, significant issues and problems for biogenesis have been identified. This is a weak hypothesis at this time.

- Conclusion: Life exists, we don't know why.

5.4 Origin of cells

Cells are very small and decompose quickly after death. As such, fossils of the earliest cells do not exist. Scientists have had to form a variety of theories on how cells (and hence life) was created on Earth.

- Bubble hypothesis
  - A. Oparin, J.B.S. Haldane, 1930's
- Primary abiogenesis: life as consequence of geochemical processes
- Protobionts: isolated collections of organic material enclosed in hydrophobic bubbles
  - Numerous variants: microspheres, protocells, protobionts, micelles, liposomes, coacervates
- Other surfaces for evolution of life
  - deep sea thermal vents
  - ice crystals
  - clay surfaces
  - tidal pools

5.5 The RNA world?

- DNA → RNA → polypeptide (protein)
- Catalytic RNA: ribozyme
  - discovered independently by Tom Cech and Sid Altman (Nobel prize)
  - catalytic properties: hydrolysis, polymerization, peptide bond formation, etc.
- Self-replicating RNA molecule may have given rise to life
  - consistent with numerous roles for RNA in cells as well as roles for ribonucleotides (ATP)
  - relationship to bubble-like structures is uncertain

5.6 The earliest cells

- Microfossils
  - ~3.5 by
  - resemble bacteria: prokaryotes
  - biochemical residues
  - stromatolites
- Archaebacteria (more properly Archaea)
  - extremophiles: salt, acid, alkali, heat, methanogens
  - may not represent most ancient life
- Eubacteria
  - cyanobacteria: photosynthesis
- atmospheric O_2; limestone deposits

- chloroplasts of eukaryotes

Cyanobacteria

5.7 Major steps in evolution of life

- Prebiotic synthesis of macromolecules
- Self replication
  - RNA? (primitive metabolism)
- DNA as hereditary material
- 1st cells
- Photosynthesis
- Aerobic respiration
- Multicellularity (more than once)

This text is based on notes very generously donated by Dr. Paul Doerder, Ph.D., of the Cleveland State University.

6 Cells

7 Cell structure

7.1 What is a cell?

The word cell comes from the Latin word "cella", meaning "small room", and it was first coined by a microscopist observing the structure of cork. The cell is the basic unit of all living things, and all organisms are composed of one or more cells. Cells are so basic and critical to the study of life, in fact, that they are often referred to as "the building blocks of life". Organisms - bacteria, amoebae and yeasts, for example - may consist of as few as one cell, while a typical human body contains about a trillion cells.

According to *Cell Theory,* first proposed by Schleiden and Schwann in 1839, all life consists of cells. The theory also states that all cells come from previously living cells, all vital functions (chemical reactions) of organisms are carried out inside of cells, and that cells contain necessary hereditary information to carry out necessary functions and replicate themselves.

All cells contain:

- Lipid bilayer boundary (plasma membrane[1])
- Cytoplasm[2]
- DNA[3] (hereditary information)
- Ribosomes[4] for protein synthesis

Eukaryotic cells also contain:

- At least one nucleus[5]
- Mitochondria[6] for cell respiration and energy

Cells may also contain:

- Lysosomes[7]
- Peroxisomes[8]
- Vacuoles[9]
- Cell walls[10]

1 http://en.wikipedia.org/wiki/plasma%20membrane
2 http://en.wikipedia.org/wiki/Cytoplasm
3 http://en.wikipedia.org/wiki/DNA
4 http://en.wikipedia.org/wiki/Ribosome
5 http://en.wikipedia.org/wiki/Cell%20Nucleus
6 http://en.wikipedia.org/wiki/Mitochondrion
7 http://en.wikipedia.org/wiki/Lysosome
8 http://en.wikipedia.org/wiki/Peroxisome
9 http://en.wikipedia.org/wiki/Vacuole
10 http://en.wikipedia.org/wiki/Cell%20Wall

7.1.1 Concepts

Plasma Membrane

Phospholipid bilayer, which contains great amount of proteins, the most important functions are the following:

1. It selectively isolates the content of the cell of the external atmosphere.
2. It regulates the interchange of substances between the cytoplasm and the environment.
3. Communicates with other cells.

Model of the fluid mosaic

Describes the structure of the plasma membrane, this model was developed in 1972 by cellular biologists J. Singer and L. Nicholson.

Phospholipid bilayer

Is in the plasma membrane and produces the fluid part of membranes.

Proteins

Long chains of amino acids.

Glucose proteins

Proteins together with carbohydrates in the plasma membrane, mostly in the outer parts of the cell.

Functions of proteins

Transport oxygen, they are components of hair and nails, and allow the cell interact with its environment.

Transport Proteins

Regulate the movement of soluble water molecules, through the plasma membrane. Some transport proteins called **channel proteins** form pores or channels in the membrane so that water soluble molecules pass.

Carrying proteins

Have union sites that can hold specific molecules.

Reception proteins

They activate **cellular responses** when specific molecules join.

Proteins of recognition

They work as identifiers and as place of union to the cellular surface.

Fluid

It is any substance that can move or change of form.

Concentration

Number of molecules in a determined unit of volume.

Gradient

Physical difference between two regions of space, in such a way that the molecules tend to move in response to the gradients.

Diffusion

Movement of the molecules in a fluid, from the regions of high concentration to those of low concentration.

Passive transport

Movement of substances in a membrane that doesn't need to use energy.

Simple diffusion

Diffusion of water, gases or molecules across the membrane.

Facilitated diffusion

Diffusion of molecules across the membranes with the participation of proteins.

Osmosis

Diffusion of the water across a membrane with differential permeability.

Transport that needs energy

Movement of substances across a membrane generally in opposition to a gradient of concentration with the requirement of energy.

Active transport

Movement of small molecules using energy (ATP).

Endocytosis

Movement of big particles towards the interior of the cell using energy. The cells enclose particles or liquids.

Pinocytosis

(Literally cell drinking) Form in which the cell introduces liquids.

Phagocytosis

Way of eating of the cells. It feeds in this case of big particles or entire microorganisms.

Pseudopods

False feet (the amoeba).

Exocitosis

Movement of materials out of the cell with the use of energy. It throws waste material.

Isotonic

The cytoplasm fluid of the interior of the cells is the same that the outer.

Hypertonic solution

The solutions that have a higher concentration of dissolved particles than the cellular cytoplasm and that therefore water of the cells goes out with osmosis.

Hypotonic

The solutions with a concentration of dissolved particles lower than the cytoplasm of a cell and that therefore do that water enters the cell with osmosis.

Swelling

Pressure of the water inside the vacuole.

Endoplasmic Reticulum

It is the place of the synthesis of the cellular membrane.

7.1.2 Structure and function of the cell

Rudolf Virchow

Zoologist, who proposed the postulates of the cellular theory, observes that the living cells could grow and be in two places at the same time, he proposed that all the cells come from other equal cells and proposed 3 postulates:

1. Every living organism is formed from one or more cells
2. The smallest organisms are unicellular and these in turn are the functional units of the multicellular organisms.
3. All the cells come from preexisting cells.

7.1.3 Common characteristics of all the cells

Molecular components

Proteins, amino acids, lipids, sweeten, DNA, RNA.

Structural components

Plasmatic membrane, citoplasm, ribosomes.

Robert Hooke

He postuled for the first time the term *cell*

Prokaryotes

Their genetic material is not enclosed in a membrane ex. Bacterias

Eukaryotes

Their genetic material is contained inside a nucleus closed by a membrane

7.2 History of cell knowledge

The optical microscope was first invented in 17th century. Shortly thereafter scientists began to examine living and dead biological tissues in order to better understand the science of life. Some of the most relevant discovery milestones of the time period include:

- The invention of the microscope[11], which allowed scientists for the first time to see biological cells
- Robert Hooke[12] in 1665 looked at cork under a microscope and described what he called cork "cells"
- Anton van Leeuwenhoek[13] called the single-celled organisms that he saw under the microscope "animalcules"
- Matthias Jakob Schleiden[14], a botanist, in 1838 determined that all *plants* consist of cells
- Theodor Schwann[15], a zoologist, in 1839 determined that all *animals* consist of cells
- Rudolf Virchow[16] proposed the theory that all cells arise from previously existing cells

In 1838, the botanist Matthias Jakob Schleiden and the physiologist Theodor Schwann discovered that both plant cells and animal cells had nuclei. Based on their observations, the two scientists conceived of the hypothesis that all living things were composed of cells. In 1839, Schwann published 'Microscopic Investigations on the Accordance in the Structure and Growth of Plants and Animals', which contained the first statement of their joint cell theory.

7.2.1 Cell Theory

Schleiden and Schwann proposed spontaneous generation as the method for cell origination, but spontaneous generation (also called abiogenesis[17]) was later disproven. Rudolf Virchow famously stated "Omnis cellula e cellula"... "All cells only arise from pre-existing cells." The parts of the theory that did not have to do with the origin of cells, however, held up to scientific scrutiny and are widely agreed upon by the scientific community today.

The generally accepted portions of the modern Cell Theory are as follows: (1) The cell is the fundamental unit of structure and function in living things. (2) All organisms are made up of one or more cells. (3) Cells arise from other cells through cellular division. (4) Cells carry genetic material passed to daughter cells during cellular division. (5) All cells are essentially the same in chemical composition. (6) Energy flow (metabolism and biochemistry) occurs within cells.

11 http://en.wikipedia.org/wiki/microscope
12 http://en.wikipedia.org/wiki/Robert%20Hooke
13 http://en.wikipedia.org/wiki/Anton%20van%20Leeuwenhoek
14 http://en.wikipedia.org/wiki/Matthias%20Jakob%20Schleiden
15 http://en.wikipedia.org/wiki/Theodor%20Schwann
16 http://en.wikipedia.org/wiki/Rudolf%20Virchow
17 http://en.wikipedia.org/wiki/Abiogenesis

7.3 Microscopes

- Allow greater resolution, can see finer detail
- Eye: resolution of ~ 100 μm
- Light microscope[18]: resolution of ~ 200 nm
- Limited to cells are larger organelles within cells
- Confocal microscopy[19]: 2 dimension view
- Electron microscope[20]: resolution of ~0.2 nm
- Laser tweezers: move cell contents

7.4 Cell size

One may wonder why all cells are so small. If being able to store nutrients is beneficial to the cell, how come there are no animals existing in nature with huge cells? Physical limitations prevent this from occurring. A cell must be able to diffuse gases and nutrients in and out of the cell. A cell's surface area does not increase as quickly as its volume, and as a result a large cell may require more input of a substance or output of a substance than it is reasonably able to perform. Worse, the distance between two points within the cell can be large enough that regions of the cell would have trouble communicating, and it takes a relatively long time for substances to travel across the cell.

That is not to say large cells don't exist. They are, once again, less efficient at exchanging materials within themselves and with their environment, but they are still functional. These cells typically have more than one copy of their genetic information, so they can manufacture proteins locally within different parts of the cell.

Key concepts: Cell size:

- Is limited by need for regions of cell to communicate
- Diffuse oxygen and other gases
- Transport of mRNA[21] and protein[22]s
  - Surface area to volume ratio limited
- Larger cells typically:
  - Have extra copies of genetic information
  - Have slower communication between parts of cell

18 http://en.wikipedia.org/wiki/Light%20microscope
19 http://en.wikipedia.org/wiki/microscopy
20 http://en.wikipedia.org/wiki/Electron%20microscope
21 http://en.wikipedia.org/wiki/RNA
22 http://en.wikipedia.org/wiki/protein

8 Structure of Eukaryotic cells

Eukaryotic[1] cells feature membrane delimited nucleii containing two or more linear chromosome[2]s; numerous membrane-bound cytoplasmic organelles: mitochondria, RER[3], SER[4], lysosomes, vacuole[5]s, chloroplast[6]s; ribosomes and a cytoskeleton[7]. Also, plants, fungi, and some protists have a cell wall.

8.1 Structure of the nucleus

The nucleus is the round object in the cell that holds the genetic information (DNA) of the cell. It is surrounded by a nuclear envelope and has a nucleolus inside.

8.1.1 Nuclear envelope

The nuclear envelope is a double-layered plasma membrane[8] like the cell membrane, although without membrane proteins. To allow some chemicals to enter the nucleus, the nuclear envelope has structures called Nuclear pore[9]s. The nuclear envelope is continuous with the endoplasmic reticulum.

8.1.2 Nucleolus

The nucleolus appears in a microscope as a small dark area within the nucleus. The nucleolus is the area where there is a high amount of DNA transcription[10] taking place.

1 http://en.wikipedia.org/wiki/Eukaryote
2 http://en.wikipedia.org/wiki/chromosome
3 http://en.wikipedia.org/wiki/RER
4 http://en.wikipedia.org/wiki/SER
5 http://en.wikipedia.org/wiki/vacuole
6 http://en.wikipedia.org/wiki/chloroplast
7 http://en.wikipedia.org/wiki/cytoskeleton
8 http://en.wikibooks.org/wiki/Plasma%20membrane
9 http://en.wikibooks.org/wiki/Nuclear%20pore
10 http://en.wikibooks.org/wiki/DNA%20transcription

8.2 Chromatin

Chromosomes consist of chromatin[11]. This is made up of strings of DNA, which typically measure centimeters in length if stretched out. This DNA is wound around a histone[12] core and organized into nucleosome[13]s.

The chromatin[14] must be uncoiled for gene expression[15] and replication[16]. Chromosome micrograph

8.3 Endoplasmic reticulum

The endoplasmic reticulum[17] is a cellular organelle[18] made up of a series of extended folded intracellular membranes. It is continuous with the nuclear membane.

There are two main types of endoplasmic reticulum:

- RER: rough endoplasmic reticulum (site of protein synthesis[19]) associated with ribosomes
- SER: smooth endoplasmic reticulum (site of lipid synthesis[20])

8.3.1 Rough Endoplasmic Reticulum

Proteins are directed to the RER by a **signal sequence** of a growing polypeptide[21]s on the ribosome. This is recognised by a signal recognition particle which brings the ribosome/polypeptide complex to a channel on the RER called a translocon. At the translocon, the signal sequence and ribosome/polypeptide complex interact with the translocon to open it. The signal sequence becomes attached to the translocon. The ribosome can continue to translate the polypeptide into the lumen of the RER. As synthesis continues, 2 processes can happen.

1. If the protein is destined to become a membrane bound protein then the protein synthesis will continue until termination. The ribosome can then dissociate, allowing protein folding within the RER lumen to occur and continuation to the golgi apparatus for processing of the polypeptide.
2. If the protein is destined for storage for later secretion after stimulation or for continuous secretion then a protease-enzyme which cuts proteins at the peptide bond-can cut the signal sequence from the growing polypeptide. Continuation to the golgi etc. can then occur.

11 http://en.wikipedia.org/wiki/chromatin
12 http://en.wikipedia.org/wiki/histone
13 http://en.wikipedia.org/wiki/nucleosome
14 http://en.wikipedia.org/wiki/chromatin
15 http://en.wikipedia.org/wiki/gene%20expression
16 http://en.wikipedia.org/wiki/replication
17 http://en.wikipedia.org/wiki/endoplasmic%20reticulum
18 http://en.wikipedia.org/wiki/organelle
19 http://en.wikipedia.org/wiki/protein%20synthesis
20 http://en.wikipedia.org/wiki/lipid%20synthesis
21 http://en.wikipedia.org/wiki/polypeptide

When produced, proteins are then exported to one of several locations. The proteins are either modified for extracellular membrane insertion or secretion. Note, this is in contrast with ribosomes which do not associate with the RER and produce proteins which will become cytosolic enzymes for example.

8.3.2 Smooth Endoplasmic Reticulum

Smooth endoplasmic reticulum produces enzyme[22]s for lipid and carbohydrate biosynthesis and detoxification RER

8.3.3 Sarcoplasmic Reticulum

This is a specialised form of endoplasmic reticulum found in some muscle cell types-particularly striated, skeletal muscle. Its main function is different from the other 2 types in that is mainly acts as a storage of calcium. This reticulum has voltage gated channels which respond to signals from 'motor neurones' to open and release calcium into the cytoplasm. This can then bring about the next part in muscle contraction.

Figure 4

Figure 1 :*Image of nucleus[23], endoplasmic reticulum and Golgi apparatus.*

1. Nucleus.
2. Nuclear pore.
3. Rough endoplasmic reticulum (RER).
4. Smooth endoplasmic reticulum (SER).
5. Ribosome on the rough ER.
6. Proteins that are transported.
7. Transport vesicle.
8. Golgi apparatus.
9. Cis face of the Golgi apparatus.
10. Trans face of the Golgi apparatus.
11. Cisternae of the Golgi apparatus.
1.

22 http://en.wikipedia.org/wiki/enzyme
23 http://en.wikipedia.org/wiki/cell%20nucleus

8.4 The Golgi apparatus

The golgi apparatus[24] is made up of multiple stacks of bilipid membranes.

- Proteins made on the RER are modified and then sorted
- Formation of secretory vesicles
- Formation of lysosomes (intracellular digestion)

Other membrane-bound cytoplasmic organelles include:

- Microbodies[25] (generic term)
- Glyoxysome (transforms fat into carbohydrate in plants)
- Peroxisome[26] (uses oxidative metabolism to form hydrogen peroxide and is destroyed by catalase[27])

8.5 Ribosomes

Ribosomes are the site of protein synthesis. Ribosomes themselves are synthesized in the cell nucleoli[28] and are structured as two subunits, the large and the small. These parts are composed of RNA and protein.

Prokaryotic and eukaryotic ribosomes are different, the eukaryotic ones being larger and more complicated.

8.6 DNA-containing organelles

Mitochondria

- Double membrane
- Aerobic metabolism, internal membrane
- DNA, ribosomes
- Give rise to new mitochondria

Chloroplast[29]

- Double membrane
- Photosynthesis, internal membrane
- DNA, ribosomes
- Give rise to new chloroplasts

Centriole[30]s

24 http://en.wikipedia.org/wiki/Golgi%20apparatus
25 http://en.wikipedia.org/wiki/Microbody
26 http://en.wikipedia.org/wiki/Peroxisome
27 http://en.wikipedia.org/wiki/catalase
28 http://en.wikipedia.org/wiki/nucleoli
29 http://en.wikipedia.org/wiki/Chloroplast
30 http://en.wikipedia.org/wiki/Centriole

- Microtubule organizing centers
  - Animal cells and many protists
  - Pair constitutes the centrosome
  - Give rise to flagellum during spermatogenesis
- Consist of 9 triplet microtubules
- Mitosis[31], meiosis[32]

8.7 Cytoskeleton

Cytoskeleton is a collective term for different filaments of proteins that can give physical shape within the cell and are responsible for the 'roads' which organelles can be carried along.

- Gives the cell shape
- Anchors other organelles
- Vital to intracellular transport of large molecules

The cytoskeleton is composed of 3 main types of filaments:

- Actin[33] filaments (7 nm)
- Microtubule[34]s: (25 nm) polymer of tubulin; 13/ring.
- Intermediate Filament[35]s

Both actin and microtubules can have associated motor proteins.

8.7.1 Intermediate Filaments

These are rope like filaments, 8-10nm in diameter and tend to give the structural stability to cells. Examples inculude Vimentin, neurofilaments and keratin. It is keratin which priniciply makes up hair, nails and horns.

8.7.2 Actin Filaments

Growth
These filaments are 2-stranded and composed of dimeric subunits called G-Actin. They contain a GTP molecule in order to bind (polymerise). As GTP is hydrolysed then the structure becomes unstable and depolymerisation occurs. The growth of actin filaments is concentration dependant-that is, the higher the concentration of free G-actin, the greater the polymerisation. The are also polar, having a + and a - end (not related to charge) and polymerisation tends to happen faster at the + end.

31 http://en.wikipedia.org/wiki/mitosis
32 http://en.wikipedia.org/wiki/meiosis
33 http://en.wikipedia.org/wiki/Actin
34 http://en.wikipedia.org/wiki/Microtubule
35 http://en.wikipedia.org/wiki/Intermediate%20Filament

Cilia and flagella are threads of microtubules that extend from the exterior of cells and used to move single celled organisms as well as move substances away from the surface of the cell. motor proteins-move, wave motion

9 Membranes

9.1 Biological membranes

Figure 5 Plasma membrane bilayer

Biological membranes surround cells and serve to keep the insides separated from the outsides. They are formed of phospholipid bilayer[1]s, which by definition are a double layer of fatty acid[2] molecules (mostly phospholipid[3]s, lipids containing lots of phosphorus).

Proteins[4] serve very important functions in cellular membranes. They are **active transports** in and out of the cell, acting as gatekeepers. They relay **signals** in and out of the cell. Proteins are the **site of many enzymatic reactions** in the cell, and play a role in **regulation of cellular processes.**

9.2 Phospholipid

Phospholipid bilayer

- basis of biological membranes and cellular organisms
- contains a charged, hydrophilic (attracted to water) head and two hydrophobic (repelled by water) hydrocarbon tails
- In presence of water, phospholipids form bilayer
  - maximize hydrogen bonds between water
  - creates barrier to passage of materials
  - fluid mosaic model shows horizontal (common) and "flip-flop" (rare) movement of phospholipids

9.3 Fluid mosaic model

- Current model of membrane
- Phospholipid bilayer
  - Phospholipids
- Move freely in lipid layer, but rarely switch layers
- Different phospholipids in each layer in different organelles
  - Glycolipids
  - Sterols (cholesterol in animals)
- Transmembrane proteins "float" in fluid lipid bilayer
  - also called intrinsic, integral proteins
- Exterior (extrinsic, peripheral) proteins

9.4 Membrane proteins

- Transport channels
- Enzymes
- Cell surface receptors

1 http://en.wikibooks.org/wiki/lipid%20bilayer
2 http://en.wikibooks.org/wiki/Fatty%20acid
3 http://en.wikibooks.org/wiki/phospholipid
4 http://en.wikibooks.org/wiki/Proteins

- Cell surface identity markers
- Cell adhesion proteins
- Attachments to cytoskeleton

Integral membrane proteins

- Anchoring to membrane
  - Protein has attached phosphatidylinositol (GPI) linkage, anchors protein in outer layer (no picture)
  - Protein has one or more hydrophobic transmembrane domains
- -helix
- -sheet

Channel protein Transport across membranes *** Diffusion**

- - From higher concentration to lower concentration
  - Membranes are selectively permeable
- Ions diffuse through membrane channels
  - Selective
  - Movement determined by diffusion and voltage differences
- **Facilitated diffusion**
  - Carrier protein, physically binds transported molecule
- **Osmosis**
- Diffusion of water down concentration gradient
- In cell: various solutes (amino acids, ions, sugars, etc.)
  - interact with water, e.g., hydration shells
- Water moves through aquaporin channels into cell
- Depends upon the concentration of all solutes in solution
  - Hyperosmotic solution: higher concentration of solutes
  - Hypoosmotic solution: lower concentration of solutes
  - Isoosmotic solution: solute concentrations equal
- Water moves from hypoosmotic solution to hyperosmotic solution

Osmotic pressure **Bulk transport**

- Endocytosis: energy requiring
  - Phagocytosis
- Solid material, typically food
  - Pinocytosis
- Primarily liquid

**** Receptor-mediated endocytosis**

- Pits on cell surface coated with clathrin and receptors
- Bind specific proteins
- Exocytosis
  - Discharge of materials from vesicle at cell surface

9.5 Receptor-mediated endocytosis

Active transport

- Energy required (usually ATP)
- Highly selective
- Works against concentration gradient
- Many examples, e.g., Na^+/K^+ pump

Cotransport (coupled transport)

- Does not use ATP directly
- Molecule is transported in connection with another molecule that is moving down a concentration gradient
  - Example: Na^+ gradient is established by a Na^+ pump, with higher concentration on outside of cell. Cotransport channel carries Na+ and another molecule (e.g. glucose) into cell
- May involve proton (H^+) pumps (chemiosmosis - ATP production)

This text is based on notes very generously donated by Dr. Paul Doerder, Ph.D., of the Cleveland State University.

10 Cell-cell interactions

with the environment with each other

10.1 Cell signaling

- Signaling requires
  - Signal
  - Cell receptor (usually on surface)
- Signaling is important in:
  - Response to environmental stimuli
  - Sex
  - Development
- Major area of research in biology today

10.1.1 Types of signaling

- Direct contact (e.g., gap junctions between cells)
- Paracrine: Diffusion of signal molecules in extracellular fluid; highly local
- Endocrine: Signal (hormone) molecule travels through circulatory system
- Synaptic: neurotransmitters

Types of signal molecules

- Hormones: chemically diverse
  - Steroid
  - Polypeptide
  - Vitamin/amino acid derived
- Cell surface proteins/glycoproteins
- Ca^{2+}, NO
- Neurotransmitter
  - Several hundred types
  - Some are also hormones e.g. Estrogen, progesterone

Receptor molecules

- Intracellular
  - Protein that binds signal molecule in cytoplasm

  - Bound receptor may act as:
- Gene regulator
- Enzyme
- Cell surface
  - Gated ion channels (neurotransmitter receptor)
  - Enzymic receptors
  - G protein-linked receptors

Cell surface protein

- Tissue identity
  - glycolipids
  - MHC proteins
- Immune systems
  - distinguish self from not-self
- Intercellular adhesion
  - permanent contact
  - help form sheets of cells, tissues
  - may permit signaling

Example: G proteins

- Transmembrane surface receptor binds signal molecule
- Conformational change allows binding of G protein on cytoplasmic side
- G protein binds GTP, becomes activated
- G protein activates intracellular signal cascade
  - Change in gene expression
  - Secrection
  - Many other possible consequences

10.2 Communicating junctions

- Gap junctions
  - animals
  - small molecules and ions may pass
- Plasmodesmata
  - plants
  - lined with plasma membrane
  - permit passage of water, sugars, etc.

10.2.1 Gap junctions

This text is based on notes very generously donated by Dr. Paul Doerder, Ph.D., of the Cleveland State University.

11 Energy and Metabolism

11.1 Energy

- The capacity to do work.
  - Kinetic energy: energy of motion (ex. jogging).
  - Potential energy: stored energy (ex. a lion that is about to leap on its prey).
- Many forms of energy: e.g.,
  - Heat
  - Sound
  - Electric current
  - Light
  - All convertible to heat
- Most energy for biological world is from sun
- Heat (energy of random molecular motion, thermal energy)
  - Convenient in biology
  - All other energy forms can be converted to heat
  - Thermodynamics: study of thermal energy
- Heat typically measured in kilocalories
  - Kcal: 1000 calories
  - 1 calorie: amount of heat required to raise the temperature of one gram of water one degree Celsius (°C)
- Heat plays major role in biological systems
  - Ecological importance
  - Biochemical reactions

11.2 Oxidation–Reduction

- Energy flows into biological world from sun
- Light energy is captured by photosynthesis
  - Light energy raises electrons to higher energy levels
  - Stored as potential energy in covalent C-H bonds of sugars
- Strength of covalent bond is measured by amount of energy required to break it
  - 98.8 kcal/mole of C-H bonds
- In chemical reaction, energy stored in covalent bonds may transfer to new bonds. When this involves transfer of electrons, it is oxidation–reduction reaction
- Always take place together
  - Electron lost by atom or molecule through oxidation is gained by another atom or molecule through reduction
  - Potential energy is transferred from one molecule to another (but never 100%)

- Often called redox reactions
  - Photosynthesis
  - Cellular Respiration
  - Chemiosynthesis
  - Autotrophs
  - Heterotrophs

11.3 NAD^+

- Common electron acceptor/donor in redox reactions
- Energetic electrons often paired with H^+

11.4 Free energy

- Energy required to break and subsequently form other chemical bonds
  - Chemical bonds: sharing of electrons, tend to hold atoms of molecule together
  - Heat, by increasing atomic motion, makes it easier to break bonds (entropy)
- Energy available to do work in a system
- In cells, G = H - TS
  - G = Gibbs' free energy
  - H = H (enthalpy) energy in molecule's chemical bonds
  - TS (T, temperature in °K; S, entropy)
- Chemical reactions break and make bonds, producing changes in energy
- Under constant conditions of temperature, pressure and volume, $\Delta G = \Delta H - T\Delta S$
- ΔG, change in free energy
  - If positive (+), H is higher, S is lower, so there is more free energy; endergonic reaction, does not proceed spontaneously; require input of energy (e.g., heat)
  - If negative (–), H is lower, S is higher. Product has less free energy; exergonic; spontaneous

===Activation energy = ==

- Reactions with $-\Delta G$ often require activation energy
  - e.g., burning of glucose
  - Must break existing bonds to get reaction started
- Catalysts lower activation energy

11.5 Enzymes

- Biological catalysts
  - Protein
  - RNA (ribozyme)
- Stabilizes temporary association between reactants (substrates) to facilitate reaction
  - Correct orientation
  - Stressing bonds of substrate

- Lower activation energy
- Not consumed (destroyed) in reaction

11.5.1 Carbonic anhydrase

- Important enzyme of red blood cells
- $CO_2 + H_2O \rightarrow H_2CO_3$ -> $HCO_3 + H^+$
- Carbonic anhydrase catalyzes 1st reaction
  - Converts water to hydroxyl
  - Orients the hydroxyl and CO_2

11.5.2 Enzyme mechanism

- One or more active sites which bind substrates (reactants)
  - Highly specific
- Binding may alter enzyme conformation, inducing better fit

11.5.3 Factors affecting enzyme activity

- Substrate concentration
- Product concentration
- Cofactor concentration
- Temperature
- pH
- Inhibitors
  - Competitive: bind to active site
  - Noncompetitive: bind to 2nd site, called allosteric site; changes enzyme conformation
- Activators
  - Bind to allosteric sites, increase enzyme activity

Cofactors

- Required by some enzymes
- Positively charged metal ions
  - e.g., ions of Zn[1], Mo, Mg, Mn
  - Draw electrons away from substrate (stress chemical bonds)
- Non-protein organic molecules (coenzymes)
  - E.g., NAD^+, $NADP^+$, etc.
  - Major role in oxidation/reduction reactions by donating or accepting electrons

1 http://en.wikibooks.org/wiki/Zinc

11.6 ATP

- Adenosine triphosphate
- Major energy currency of cells, power endergonic reactions
- Stores energy in phosphate bonds
  - Highly negative charges, repel each other
- Makes these covalent bonds unstable
  - Low activation energy
- When bonds break, energy is transferred
- ATP $\rightarrow$ ADP + Pi + 7.3 kcal/mole

11.7 Biochemical pathways

- Metabolism: sum of chemical reactions in cell/organism
- Many anabolic and catabolic reactions occur in sequences (biochemical pathways)
- Often highly regulated

Evolution of biochemical pathways

- Protobionts or 1st cells likely used energy rich substrates from environment
- Upon depletion of a substrate, selection would favor catalyst which converts another molecule into the depleted molecule
- By iteration, pathway evolved backward

This text is based on notes very generously donated by Paul Doerder, Ph.D., of the Cleveland State University.

12 Respiration: harvesting of energy

Glucose + $O_2 \rightarrow CO_2 + H_2O$ + ATP

12.1 Energy

- Energy is primarily in C-H bonds (C-O too)
- Chemical energy drives metabolism
  - Autotrophs: harvest energy through photosynthesis or related process (plants, algae, some bacteria)
  - Heterotrophs: live on energy produced by autotrophs (most bacteria and protists, fungi, animals)
- Digestion: enzymatic breakdown of polymers into monomers
- Catabolism: enzymatic harvesting of energy
- Respiration: harvesting of high energy electrons from glucose

12.2 Respiration

- Transfer of energy from high energy electrons of glucose to ATP
- Energy depleted electron (with associated H^+) is donated to acceptor molecule
  - Aerobic respiration: oxygen accepts electrons, forms water
  - Anaerobic respiration: inorganic molecule accepts hydrogen/electron
  - Fermentation: organic molecule accepts hydrogen/electron

12.3 Respiration of glucose

- $C_6H_{12}O_6 + 6\ O_2 \rightarrow 6\ CO_2 + 6\ H_2O$ + energy
- ΔG = -720 kcal/mole under cellular conditions
- Largely from the 6 C-H bonds
- Same energy whether burned or catabolized
- In cells, some energy produces heat, most is transferred to ATP

12.4 Alternative anaerobic respiration

- Methanogens (Archaebacteria).
  - CO_2 is electron acceptor, forming CH_4

- Sulfur bacteria
  - SO_4 reduced to H_2S
  - Formation of H_2S set stage for evolution of photosynthesis (H_2S as electron donor before H_2O)
  - About 2.7 by, based on ratio of 32S/34S, where only biological processes produce 32S enrichment

12.5 Glycolysis overview

Glycolysis accounting

- Oxidation
  - Two electrons (one proton) are transferred from each G3P to NAD^+ forming NADH

2NADH

- Substrate level phosphorylation
  - G3P to pyruvate forms 2 ATP molecules

```
4 ATP (from 2 G3P)
```

–2 ATP (priming)

```
2 ATP (net gain)
```

Summary: The net input of glycolysis is 2 ATP molecules which are used to split one glucose molecule. The net yield of this step is 2 ATP and 2 pyruvate.

12.6 Regeneration of NAD^+

- Reduction of NAD^+ to NADH can deplete NAD^+ supply; it must be regenerated
- Two pathways, coupled to fate of pyruvate
  - With oxygen: enter electron transport chain, forming water (and ATP)
  - Without oxygen: fermentation
- lactate
- ethanol

12.7 Alcohol fermentation

12.8 Lactate formation

Either lactic acid[1] or alcohol can be formed as a result of anaerobic respiration in cells.

1 http://en.wikipedia.org/wiki/Lactate

12.9 Krebs cycle: overview

- Matrix of mitochondrion
- Priming steps
  - Joining of acetyl-CoA to oxaloacetate
  - Isomerization reactions
- Energy extraction steps in Krebs cycle
  - Per glucose
- 6 NADH
- 2 $FADH_2$
- 2 ATP (from GTP)
- 4 CO_2

12.10 ATP production

- Chemiosmosis (Mitchell)
- H^+ (from NADH and $FADH_2$) is pumped against a gradient into the intermembranal space of the mitochondrion (creates voltage potential)
- Diffusion back into matrix through ATP synthase channels drives synthesis of ATP (ADP + Pi $\rightarrow$ ATP)
- ATP exits mitochondrion by facilitated transport

12.11 Evolution of aerobic respiration

- Preceded by evolution of photosynthesis (O2 needed; also, prior evolution of electron transport and chemiosmosis)
- High efficiency of ATP production compared to glycolysis
  - Fostered evolution of heterotrophs
  - Fostered evolution of mitochondria by endosymbiosis in eukaryotes

13 Photosynthesis

$$\mathbf{6\ CO_2 + 6\ H_2O \rightarrow C_6H_{12}O_6 + 6\ O_2}$$

- One of most important reactions in history of life:
  - source of atmospheric O_2
  - ultimately led to aerobic respiration and eukaryotes
- Responsible for bulk of glucose production
- Early experiments showed that mass of plant must be derived from substances in the air, not the soil
- Experiments with isotopes showed that liberated oxygen comes from water
- Experiments also showed that light is essential but that some reactions (e.g., reduction of CO_2) continue in the dark
- Plants do two big, important things during photosynthesis: gain energy (absorb light) and build sugar (glucose).
- Photosynthesis can be divided into two series of chemical reactions: the light (light-dependent) reactions and the dark (light-independent) reactions. In light reactions, light is absorbed; in dark reactions, sugar is built.
- Occurs when plants, algae, and autotrophic bacteria absorb light energy and build glucose.

13.1 Light Reactions

- Part of the electromagnetic spectrum
- Consists of units of energy called photons
- Photons at UV end of spectrum have more energy than those at the red end
- Occur on the surface of thylakoid disks
- Chlorophyll and other plant pigments differentially absorb photons
  - Chlorophyll a: light to chemical energy
  - Chlorophyll b: accessory chlorophyll
  - Chlorophylls absorb primarily blue and red (green reflected back, hence the green color of plants)

13.1.1 Accessory pigments

- Chlorophyll is a major light gathering pigment
  - Absorbs light with considerable efficiency (i.e., retaining energy)
- Accessory pigments
  - Chlorophyll b
  - Carotenoinds
- capture light of wavelengths not captured by chlorophylls

- Confer other colors to plant leaves (autumn colors too)

Photosynthetic steps

- Primary photoevent: light photon captured by photosystem and energy transferred to electron donated by water
- Electron transport: excited electron is shuttled along imbedded series of electron carriers to proton pump and electron is transferred to acceptor
- Chemiosmosis: transport of protons back into chloroplast drives synthesis of ATP

13.1.2 The Even More Detailed Light Reactions

What the Light Reactions Do:

The light reactions of photosynthesis occur in chloroplasts in and on the thylakoid disks. During the light reactions, light energy charges up ATP molecules. More specifically, light turns the chloroplast into an acid battery, and this battery charges up ATP.

How the "Chloroplast-Battery" Charges ATP:

The **stroma** is the fluid inside of the chloroplasts, and it carries a negative charge. This means that it contains about a "gazillion" extra electrons. The solvent of stroma is water.

The fluid inside the thylakoid disks is positively charged because it contains a lot of hydrogen (H+) ions. The pH here is low, making the fluid very acidic. The solvent of thylakoid disk fluid is water.

A chloroplast acts like a battery, because it has separated a strong positive charge and a strong negative charge in two different compartments. Energy is released when H+ ions (free protons) flow from the inside of a thylakoid disk to the stroma. This is electrical energy, since it is a flow of charged particles.

The protons pass through special channels (made of protein) in the thylakoid membrane; this reaction is 'exothermic.' The energy that is given off is used to fuel this reaction (Pi is the phosphate ion):

```
ADP + Pi --> ATP
```

The proton can go to the negative stroma, but only if it uses its energy to charge up ATP. Since one reaction wants to go, and the other one doesn't, and since the first reaction releases energy and the second one absorbs energy, the two reactions are known to be 'coupled' together so that the first fuels the second. Of course, a special enzyme must be involved for this to happen.

Chlorophyll Molecules on a Thylakoid Disk:

Hundreds of chlorophyll molecules cover the surface of a thylakoid disk, making the disk green. The nonpolar "tails" of the chlorophyll molecule are embedded in the membrane of the thylakoid.

13.2 “Dark” reactions

- ATP drives endergonic reactions
- NADPH provides hydrogens for reduction of CO_2 to carbohydrate (C-H bonds)
- Occur in the stroma
- First step in carbon fixation

13.2.1 The Detailed Dark Reactions

What the Dark Reactions Do:

The dark reactions build sugar from carbon dioxide gas (CO2), water (H2O), and energy from ATP molecules that were charged up during the light reactions. The dark reactions occur in the stroma of a chloroplast. Dark reactions usually occur in the light, but they don't have to. They'll occur in the dark until the chloroplast's supply of ATP runs out (usually about 30 seconds).

The Calvin Cycle:

The Calvin Cycle is the fancy name for the metabolic pathway that builds sugar. This means that it involves a whole lot of chemical reactions, and it uses a lot of different enzymes to catalyze the reactions.

Carbon dioxide gas is stable, therefore the bonds that hold the carbon and oxygen atoms are strong. Therefore it takes a lot of energy to break the bonds and separate the carbon atoms from the oxygen atoms. The energy needed to do this comes from ATP molecules.

When inorganic carbon (like from CO2) is being added to an organic molecule (such as sugar), this is called carbon fixation.

It takes 2 complete turns of the Calvin Cycle to make a glucose molecule.

Some portions of this text is based on notes very generously donated by Paul Doerder, Ph.D., of the Cleveland State University. The detailed portions are not provided by Dr. Doerder.

How cells divide

13.3 Prokaryote cell division

- Binary fission
  - Doubling of cell contents, including DNA
  - Fission to divide contents
- Segregation of replicated genomes by growth of membrane between attachment points
- Partitioning of cytoplasmic components
- Escherichia coli
  - Capable of cell division every 20 minutes under optimal conditions (DNA in continuous state of replication)
  - Model organism of bacterial cell division

13.4 Bacterial DNA replication

- Replication follows rules of base pairing, with each polynucleotide chain serving as template for synthesis of its complement.
- Genetic evidence showed that the bacterial chromosome is circular long before there was corroborating physical evidence.

Eukaryotic chromosomes

- Discovered by Walther Fleming in 1882 in dividing cells of salamander larvae, following improvements in microscopes and staining technology
  - He called division mitosis (mitos = "thread")
- Chromosome number is constant in a species
  - Ranges from 2 to >500 (46 in human somatic cells)
  - Homologous pairs, one contributed by each parent
  - Change in number is cause and consequence of speciation
- Chromosome constancy and their precise division in mitosis and meiosis led biologists to postulate that they were carriers of hereditary information

13.5 Chromosome number

- 1N = number of chromosomes in gamete
- 1N = haploid chromosome number
- 2N = number of chromosomes in somatic cells (cells that are not egg or sperm)
- 2N = diploid
- Deviations from N or 2N are usually lethal in animals

Chromosome numbers

13.6 Eukaryotic chromosomes

- Consist of chromatin
  - DNA and associated proteins, mainly histones
  - Nucleosomal organization
  - Euchromatin: unwound chromatin, in basic nucleosomal configuration; genes available for expression
  - Heterochromatin: highly condensed except during replication
- Karyotype: array of chromosomes an individual possesses
  - Clinical importance (Down syndrome; cancer)
  - Evolutionary importance (speciation)

13.7 Chromosome organization

13.8 Human karyotype stained by chromosome painting

13.9 Chromosomes

- Homologous pairs
  - Inherited one from each parent
  - Identical in length and position of centromere
  - Contain identical or similar genes
  - Homologous pair = homologs
- Morphology
  - After replication, consist of two sister chromatids attached to a centromere

13.10 Human chromosomes

- Diploid number = 2n = 46 = 23 pairs of homologs
- Haploid number = 23 (gametes)
- Each replicated chromosome contains 2 sister chromatids = 92 chromatids

Cell cycle

- Growth and division cycle of cells
- Precisely controlled by biochemical and gene activity, except in cancer
- Phases
  - G1: primary growth phase
  - S: DNA replication; chromosome replication
  - G2: second growth phase; preparation for mitosis
  - M: mitosis; nuclear division
  - C: cytoplasmic division

13.11 Mitotic cell cycle

- Cells exiting the cell cycle are said to be in G0
- Cell cycle time varies with stages of life cycle and development, with G1 the most variable
- DNA replication occurs during S phase of the cell cycle following G1.

```
 - at this point the chromosomes are composed of two sister
chromotids connected by a common centromere.
```

13.12 Replicated human chromosomes

13.13 Mitosis

- Nuclear division
  - equational division of replicated chromosomes
  - chromatids move to opposite poles
- Continuous process
  - prophase
  - metaphase
  - anaphase
  - telophase
- Driven by motors and microtubules
- No change in chromosome number
  - N → N by mitosis
  - 2N → 2N by mitosis
- May be accompanied by cytokinesis

Kinetochore Microtubules attach to kinetochores. Metaphase

- Momentary alignment of chromosomes in center of cell

Anaphase

13.14 Plant mitosis

- Similar to animal mitosis
- New cell wall formed between cells from membrane partition

Cell cycle control

- Cell cycle events are regulated by protein complexes and checkpoints
- Discovered by microinjection of proteins in to eggs, by mutational analysis and by techniques of molecular biology

Molecular control of cell cycle: Cdk and cyclin

- Cyclin dependent protein kinase (Cdk)
  - Phosphorylate serine/threonine of target regulatory proteins
  - Function only when bound to cyclin
- Cyclin: short-lived proteins that bind to cdks

13.15 Controlling the cell cycle

- External signals initiate cell division in multicellular organisms
- Growth factors: extracellular regulatory signals
  - Usually soluble; bind to cell surface receptor

- Sometimes membrane bound, requiring cell-cell contact with receptor
- E.g., upon wound, platelets release PDGF which stimulates fibroblasts to enter cell cycle (exit G0), to heal wound

13.16 Cancer

- Unregulated cell proliferation
- Cancer cells have numerous abnormalities
  - >46 chromosomes
  - Mutations in proto-oncogenes
- Encode proteins stimulating the cell cycle
- May be regulated by phosphorylation
- Often over expressed in cancer cells
  - Mutations in tumor-suppressor genes
- Encode proteins inhibiting the cell cycle
  - Often bind to products of proto-oncogenes
- May be regulated by phosphorylation

13.17 Mutations and cancer

This text is based on notes very generously donated by Paul Doerder, Ph.D., of the Cleveland State University.

14 Sexual reproduction

14.1 Sexual

- Exclusively eukaryotes
- Fusion of two haploid genomes
  - Fertilization (= syngamy)
  - Forms new individuals in multicellular organisms as result of fusion of egg and sperm
- Plants
- Animals
- Meiosis yields haploid genomes at some point in life cycle

14.2 Sexual life cycle

Typical animal life cycle

- Meiosis occurs in germ line cells in gonads producing haploid gametes
- All other cells are somatic cells
- Alternation of generations
- Sexual intercourse

14.3 Meiosis

- Gives rise to genetic variation
- Reduction division: 2n to n
- Preceded by one round of DNA (chromosome) replication
- Two rounds of nuclear (& usually cell) division
  - Meiosis I
- Synapsis of homologs
- Segregation of homologs
- Reduction division, 2n to n
  - Meiosis II
- No chromosome replication
- Segregation of sister chromatids
- Formation of 4 haploid (n) cells
- Two nuclear divisions, usually 2 cell divisions, only one round of replication
- Meiosis I
  - Prophase: synapsis and crossing over
  - Metaphase

  - Anaphase: chromosome segregation
  - Telophase
- Meiosis II (mitosis-like)
  - Prophase
  - Metaphase
  - Anaphase: sister chromosome segregation
  - Telophase

14.4 Prophase I: synapsis

- Complete alignment of replicated homologs
- Synapsis occurs throughout the entire length of a pair of homologs
- Key to chromosome segregation
- Synapsis, crossing over
- Subdivided into 5 continuous stages

14.5 Crossing over

- Reciprocal, physical exchange between nonsister chromatids
- Type of recombination; mixes maternal and paternal genes
- Visual evidence: chiasmata

14.6 Microtubules and anaphase I

- During prophase microtubules attach to kinetochores on one side of centromere
- The metaphase checkpoint insures proper attachment
- A phosphorylation event initiates motor activity and anaphase

14.7 Meiosis II

- Cytologically similar to mitosis
  - No preceding DNA replication
  - Chromatids segregate and move to opposite poles as chromosomes
  - 4 haploid cells produced
- In animals, these cells differentiate into gametes
- In plants and many other organisms, these cells divide by mitosis, followed some time later by gamete formation

14.8 Evolution of sex

- Asexual reproduction: all offspring genetically identical to parent

- Sex: recombination destroys advantageous combinations
- So why sex?
  - Many hypotheses
  - Effect repair of genetic damage?
- Much pachytene repair as well as gene conversion
- Some protists form diploid cells in response to stress
  - Recombination breaks up combinations of genes favoring parasites, thus reducing parasitism?

14.9 Consequences of sex

- Recombination: generates genetic diversity
  - Crossing over
  - Independent assortment
- Random fertilization
  - Qualities of gamete usually do NOT reflect qualities of genes enclosed in gamete

This text is based on notes very generously donated by Paul Doerder, Ph.D., of the Cleveland State University.

15 Genetics

16 Gregor Mendel and biological inheritance

Charles Darwin[1], for all he contributed to the science of biology, never knew about the mechanism by which living things inherit traits from previous generations, or how new traits arise.

As any schoolchild can tell you, this mechanism of interitance has since been found to be **DNA**[2], or deoxyribonucleic acid. DNA allows for stable inheritance of traits: the code in each strand of DNA is replicated precisely through the pairing of basic units along each strand. The error rate in this replication is amazingly low; not even one base pair in a million matches out of sequence.

However, when even one base pair is added to a new strain of DNA in an order differently than in the parent chain, it can be the basis of a **mutation**. These changes in DNA sequences are the microscopic origin of changes in traits of all studied living things. Even the smallest difference in a strand of DNA can result in a change in traits that can cost the life of the organism. Mutations can produce proteins with a new or altered function. In humans, the example of Sickle cell anemia[3] is commonly given as its origin is a difference of only one base pair in a section of DNA that encodes red blood cells.

Individual sequences of DNA that encode for specific proteins are called **genes** and are the units of heredity. Each one has a set nucleotide[4], and together all of the genes (and some sequence of DNA that apparently do not code for any biologically important functions) together make up the entire **chromosome**[5]

16.1 Mendel

- Discovered principle of genetic segregation via numerous experiments utilizing pea plants
- Inferred the existence of genes through segregation of phenotypes
- Used quantitative methods: counted; ratios
- Work is model of scientific method
- In particular, observed the F2 progeny, which lead to the discovery of dominant and recessive traits
- Published work in 1866, went unnoticed
- In 1900 his scientific paper was "rediscovered"

1 http://en.wikibooks.org/wiki/Charles%20Darwin
2 http://en.wikipedia.org/wiki/DNA
3 http://en.wikipedia.org/wiki/Sickle%20cell%20anaemia
4 http://en.wikipedia.org/wiki/nucleotide
5 http://en.wikipedia.org/wiki/chromosome

- Mendel is acknowledged as founder of Genetics
  - still used alphabet letters to designate genes
  - still refer to dominant and recessive genes
  - still refer to segregation of alleles in meiosis
  - principle of segregation applies to all sexually reproducing organisms; Mendel's results were immediately applied to humans in 1900

16.2 Mendel's experiments

- 1856, began experiments with the garden pea, Pisum sativum
- 1865, presented results to the Bruno natural history society, which he helped found
- 1866, published his results in proceedings of the society
- Naegeli encouraged Mendel to reproduce results in another species, which failed because the species did not undergo true fertilization
- discrete traits in Pisum sativum
  - pure-breeding lines
  - dominant/recessive alleles
- alleles are two alternate versions of a gene
  - gametes contain hybridized chromosomes that are formed during meiosis
  - homozygous has two of the same allele
  - heterozygous has two different alleles
- reciprocal F1 crosses (all exhibiting dominant phenotypes); F2; F3
- counted offspring, noted ratios
- inferred genotypes from phenotypes
- tested hypotheses with testcrosses
- attempted to repeat with another species

16.3 Mendel's seven pairs of traits

1. Seed form (round or wrinkled)
2. Cotyledon color (green or yellow)
3. Seed coat color (white or colored)
4. Pod form (inflated or constricted)
5. Pod color (green or yellow)
6. Flower position (axial or terminal)
7. Plant heights (tall or short)

16.4 Locus

- The location of a specific gene within a chromosome

16.5 Modern Y chromosome

Y-chromosome is the most evolved chromosome. Generally it is thought that if Y- chromosome is present in an individual then he will be male. But if mutation occurs at sex determining region or zinc factor then it will not code for testis determining factor, and results in normal female. This type of female's frequency is 1/250000.

16.6 Chromosome phenomena

- X-chromosome inactivation
  - Barr bodies
- Nondisjunction: failure of chromosome segregation at meiosis or mitosis
  - Results in 2N ± 1 chromosome number
- Trisomy 2N + 1
  - Usually lethal. Trisomy 21 (Down) exception
- Monosomy 2N ** 1
  - Lethal except XO
  - Usually maternal origin in humans

16.7 X-chromosome inactivation

In females, one X-chromosone is randomly switched off forming a Barr body.

16.8 Barr body

Dense region in the nucleus formed by the inactive X-chromosome.

16.9 Human genetic disorders

Down's Syndrome(Mongolism)

Down's Syndrome is usually produced by the nondisjunction of chromosome 21 during oogenesis and sometimes during spermatogenesis. The individual suffering from this type of syndrome has 47 chromosomes instead of the normal 46. The extra chromosome is not a sex chromosome but an autosome.

Most cases of mongolism were found to occur in children born by women in their forties. The affected children, called mongoloids, show mental retardation and have a shorter life expectancy. Their most prominent feature is the Mongolian folds in their eyes; hence, the term mongolism.

Klinefelter's Syndrome

When an XY-bearing sperm unites with an X-bearing egg, the resulting condition is called Klinefelter's Syndrome, or sexually undeveloped male. Individuals having the syndrome show the following characteristics:

- testes are small
- sperms are never produced
- breasts are enlarged
- body hair is sparse
- individuals are mentally defective

The same abnormal meiotic division may occur in females. They produce eggs with XX or no sex chromosomes. Such egg, when fertilized by a Y-bearing sperm, will not develop (YO). This is because YO is lethal--it wil cause death to the offspring.

17 DNA: The Genetic Material

17.1 DNA

DNA stands for *Deoxyribose Nucleic Acid*. That is, a nucleic acid with two sugars. DNA is the hereditary material of cells and is considered the blueprint of life. DNA is found in all kingdoms of life. Even most viruses have DNA. A molecule of DNA is chemically stable (it does not have a 2-prime alcohol group.)

When someone says DNA, they may be referring to one's genetic material on multiple levels: They may be speaking about a single deoxyribose nucleic acid molecule, a section of a double helix, a section of a chromosome, or one's entire hereditary composition.

- antiparallel
- Double helix
  - Semiconservative replication
  - Sequence of nucleotides encodes functional RNA or polypeptide

17.2 Historical perspective

- Mitosis and meiosis
  - Regular distribution of chromosomes suggested that they contain hereditary information
  - Bridges/Morgan, using Drosophila melanogaster showed that genes are on chromosomes (1910s)
- Hammerling: nucleus contains hereditary information (1930s)
- Griffith: transformation of bacteria (1928)
- Avery, MacLeod, McCarty: transforming substance is DNA (1944)
- Hershey, Chase: DNA is hereditary material of viruses (1952)
- Rosalind Franklin
- Watson and Crick: structure of DNA (1953)

17.3 Hershey-Chase Experiment

The Hershey and Chase experiment was one of the leading suggestions that DNA was a genetic material. Hershey and Chase used phages, or viruses, to implant their own DNA into a bacterium. They did two experiments marking either the DNA in the phage with a radioactive phophorus or the protein of the phage with radioactive sulfur. With the bacteria that was infected by the phages with radioactive DNA the DNA in the bacteria

was radioactive. In the bacteria that was infected with the radioactive protein the bacteria was radioactive, not the DNA. This proves that DNA is a genetic material and it is passed on in viruses.

17.4 DNA/RNA components

- Miescher: discovered DNA, 1869

17.4.1 Structure of DNA

DNA is in a double helix structure made up of nucleotides. The "backbone" of the double helix is composed of phosphates connected to a five carbon sugar called *deoxyribose*, . The "rungs" are composed of nitrogenous bases, Purines and Pyrimidines. Purines contain Adenine(A) and Guanine(G) and have two rings in their structures. Pyrimidines contain Cytosine(C) and Thymine (T) and have one ring in their structures.

17.5 Chemical structure of DNA

- Polynucleotide
  - Phosphodiester bonds between nucleotides
  - 5'-pGpTpCpGpTpApApTp-OH 3'
- Chargaff's rules, in DNA: equimolar amounts
  - A = T
  - G = C

17.6 3D structure of DNA

- James Watson[1] and Francis Crick[2] (1953[3])
  - Nucleotide
  - Keto and amino forms of bases
  - Chargaff's rules
  - X-ray crystallographic data (Rosalind Franklin[4])

17.7 Franklin

- X-ray diffraction of DNA crystals
- revealed regular pattern explained by antiparallel double helix

1 http://en.wikipedia.org/wiki/James_D._Watson
2 http://en.wikipedia.org/wiki/Francis_Crick
3 http://en.wikipedia.org/wiki/1953
4 http://en.wikipedia.org/wiki/Rosalind_Franklin

DNA model

- Double helix of polynucleotides
  - antiparallel
  - 3'-5' phosphodiester bonds
- Base pairs held by hydrogen bonds
  - AT
  - GC
- There are about 10 base pairs per turn of helix
- model has predictive power
  - mode of DNA replication
  - encoding of genetic information

17.8 DNA replication

- Conservative model
  - One double helix of both old strands
  - One double helix of two new strands
- Dispersed
  - Each strand mixture of old new
- Semiconservative
  - Meselson-Stahl experiment confirmed its viability over the previous two
    - grew *E. coli* bacterium in a culture containing 15N (a heavy isotope of nitrogen)
    - bacterium assimilated the 15N into their DNA
    - a similar process was then done using 14N, a lighter isotope
    - following centrifugation, the densities were observed to be that of combined in the middle, and 14N on top, thereby confirming the semiconservative model

17.9 DNA replication

- Semiconservative
- New nucleotides added to 3' –OH
- Replication fork
  - Replication complex
- DNA polymerase
- Associated enzymes/proteins
  - Energy from phosphate bonds of triphosphate nucleotide substrates (dNTP)

17.10 DNA polymerases

- Prokaryotes, E. coli
  - 3 DNA polymerases
  - III is main enzyme for DNA replication
  - ~1000 nt/sec

- Eukaryotes
  - 6 DNA polymerases
- Add nucleotide to 3' –OH end
- All require primer, i.e., free 3' –OH

17.11 DNA replication complex

- Helicase "unzips" the DNA double helix
- Primase: synthesize RNA primer
- Single-strand binding proteins
- DNA gyrase (topoisomerase)

- DNA polymerase III
- DNA polymerase I (remove primer, fill gaps)

17.12 DNA replication

- 5' → 3' replication
  - Nucleotide addition at 3' –OH
  - No exceptions
- New strands are oriented in opposite direction due to 5' → 3' constraint
  - Leading strand: continuous replication
  - Lagging strand: discontinuous replication
    - contains multiple Okazaki fragments
- Joined by DNA ligase

17.13 DNA replication fork

- primer required by all DNA polymerases

17.14 Replication units

17.15 Replicon

A region of DNA that is replicated from a single origin.

17.16 What is gene?

- Garrod
  - "inborn errors of metabolism"
  - Alkaptonuria: enzyme deficiency

- Beadle and Tatum
  - One gene one enzyme
  - Genetic and biochemical analysis in Neurospora
- Today: gene is sequence of nucleotides encoding functional RNA molecule or the amino acid sequence of a polypeptide

This text is based on notes very generously donated by Paul Doerder, Ph.D., of the Cleveland State University.

18 Gene expression

Flow of genetic information

- DNA → mRNA → polypeptide
- Transcription: DNA → mRNA
  - RNA polymerase
  - Nucleus in eukaryotes
  - Transcription also makes rRNA and tRNA
- Translation: mRNA → polypeptide
  - Ribosomes: protein and rRNA
  - Genetic code and tRNA

18.1 "Central Dogma"

18.2 The Genetic Code

- Triplet codon
  - 64 triplet codons (43)
  - Experimentally deciphered in 1961
- Nearly universal
  - Implies common ancestor to all living things
  - Minor exceptions: certain ciliates, mitochondria, chloroplasts
- Still evolving

18.3 Transcription

- RNA polymerase
  - NTP substrates
  - Synthesizes single stranded RNA complementary to template strand of DNA
  - New nucleotides to 3' end
- Begins at promoter site
  - no primer necessary
- Ends at terminator site
- Much posttranscriptional modification in eukaryotes

18.4 Transcription bubble

Promoter site

- Prokaryotes
  - -10 nt, TATA box
  - -35 nt, additional signal
- Eukayotes
  - -25, TATAAA box
  - Additional signals upstream
- Promoters may be strong or weak
- In eukaryotes, access to promoter depends upon state of chromatin coiling

18.5 Eukaryote mRNA

- Synthesized as pre-mRNA, processed in nucleus
- 5' end: GTP cap placed in inverted position
  - Essential for ribosome recognition
- 3' end: poly-A tail; non-templated addition of ~50-250 A nucleotides; stability
- Introns: intervening sequences removed

18.6 Translation

- Requires:
  - mRNA
  - tRNA
  - ribosomes
  - translation factors (various proteins)
- In prokaryotes, takes place on growing mRNA
- In eukaryotes, in cytoplasm on free ribosomes and RER
- AUG start codon to stop codon

18.7 Translation in bacteria

tRNA

- Transfer RNA
- Two important parts
  - Anticodon
- Hydrogen bonds with mRNA codon
  - 3' end
- Accepts amino acid (using energy of ATP)
- Aminoacyl-tRNA synthetase

18.8 Aminoacyl tRNA synthase

- Enzyme used to bind amino acid from the cytoplasm to tRNA, which then transfers that amino acid to the ribosome for polypeptide formation

18.9 Ribosome structure

18.10 Large ribosome subunit

18.11 Translation

- Initiation complex
  - Small ribosomal subunit
  - mRNA
  - fMet-tRNA (prokaryotes only; met-tRNA in eukaryotes)
  - Initiation factors
- Elongation
  - Ribosome
  - mRNA
  - tRNAs
  - Elongation factors

18.12 Initiation complex

18.13 Elongation, translocation

- incoming tRNA enters the A site
- rRNA catalyzes peptide bond formation. Note that growing peptide is attached to what was incoming tRNA at P site after translocation.
- empty tRNA leaves via E site; recycled
- A site ready for next charged tRNA

18.14 Introns/exons

- In eukaryotes, coding regions of gene may be interrupted by introns, noncoding regions of DNA (RNA)
- Introns
  - 22- >10,000 nt in length
  - 5' GU … 3' AG removal sequence
  - Not essential to genes
  - May constitute >90% of gene

- removed from pre-mRNA to form mRNA
- Exon: often codes for functional domain of protein
  - translatable mRNA

This text is based on notes very generously donated by Paul Doerder, Ph.D., of the Cleveland State University.

19 Gene regulation

- Not all genes are expressed in a cell
- Gene expression can be turned on and off
- Multiple levels of regulation gene function
  - Transcription initiation
- State of chromatin
- Transcription factors
  - Post-transcriptional
- mRNA processing
- mRNA half-life
- Translational
- Post-translational
  - Protein modification

19.1 Transcriptional control

- State of chromatin
  - Euchromatin: transcriptionally active
  - Heterochromatin: transcriptionally inactive
  - Chemical modification of histones
  - Methylation of bases
- Transcription factors
  - Bind to DNA at promoter or other regulatory sites (enhancers)
- Recognize base sequence through major and minor grooves
  - Recruit RNA polymerase

19.2 DNA grooves

Categories of transcription factors in eukaryotes

- Helix-turn-helix
  - Two small "-helices
  - Fit into DNA groove
- Homeodomain
  - Highly conserved helical domains
  - ˜60 amino acids
- Zinc finger motif
  - Zn atom bound
- Leucine zipper

- dimer

19.3 Regulatory proteins

- Activity may depend upon allosteric binding of small molecules
  - cAMP
  - Co-repressors
  - Inhibitors
- Binding to promoter region may "bend" DNA, making it accessible to other regulatory proteins

19.4 Lac operon of E. coli

- Single promoter region for cluster of genes
- Regulated and transcribed as a single unit
- Operons typical in prokaryotes
- Repressor: turns OFF gene expression

lac repressor

- Turns off transcription by blocking access by RNA polymerase
- repressor in activated by allosteric binding of lactose

Regulation in eukaryotes

- Both proximal (promoter) and distal (enhancer) to gene
- Typically transcription unit encodes a single polypeptide
- Promoter
  - TATA box
  - Other elements (regulatory sequences) may be present
- Enhancers
  - Work upstream, downstream, close, far from gene
  - Bend DNA

19.5 Alternative splicing

- Single transcript gives rise to 2 or more mature mRNAs
  - encode different polypeptides with shared domains
  - tissue and developmentally specific

This text is based on notes very generously donated by Paul Doerder, Ph.D., of the Cleveland State University.

(This Page was Last Edited December 2005)

20 Mutation

A mutation is a permanent change to an organism's genetic material (DNA[1] or RNA). Mutations are a rare but significant biological process, since they provide the variation on which evolution[2] acts and are also the source of cancer.

An organism's genetic material is made up of polymers (chains) of four different nucleotides[3], like a recipe book written in a language of only four letters. A mutation event is when the order of the nucleotides in DNA change, usually when the DNA is being copied.

Mutations come in a number of forms:

20.1 Point Mutations

Point mutations are all mutations which involve a single nucleotide. These come in the form of substitutions, insertions and deletions:

20.2 Substitution

Substitution Mutations: In substitution mutations, a nitrogenous base of a triplet codon of DNA is replaced by another nitrogen base or some derivative of the nitrogen base, changing the codon. The altered codon codes for a different amino acid substitution.The substitution mutations are of two types:

1.Transitions: It is the replacement of one purine in a polynucleotide chain by another purine(A by G or C by A) or one pyrimidine by another pyrimidine(T by C or C by T)

2.Transversions:A base pair substitution involving the substitution of a purine by pyrimidine or pyrimidine by a purine is called transversion.

1 http://en.wikipedia.org/wiki/DNA
2 http://en.wikibooks.org/wiki/General%20Biology%2FEvolution
3 http://en.wikipedia.org/wiki/nucleotide

20.2.1 Insertion

20.2.2 Deletion

20.3 Larger mutations

Larger mutations which involve more than one nucleotide also include insertions and deletions, but can also include inversions, rearrangement of nucleotides and duplication of entire genes:

20.3.1 Inversion

20.3.2 Rearrangement

20.3.3 Gene/Exon Duplications

Transposition

Retrotransposition

20.4 Chromosomal mutations

Chromosomal mutations involve changes to entire chromosomes. These mutations are particularly rare:

20.4.1 Translocation

20.4.2 Fusion

20.4.3 Fission

20.4.4 Segmental Duplication

20.4.5 Chromosomal Duplication

20.4.6 Genome Duplication

20.5 Causes of mutations

20.6 Effects of mutations

Mutations can have a variety of different effects depending on the type of mutation, the significance of the piece of genetic material affected and whether the cells affected are germ-

line cells. Only mutations in germ-line cells can be passed on to children, while mutations elsewhere can cause cell-death or cancer.

Mutations can be classified by their effects:

20.6.1 Silent Mutation

Silent Mutations are DNA mutations that do not result in a change to the amino acid sequence or a protein. They may occur in a non-coding region (outside of a gene or within an intron), or they may occur within an exon in a manner that does not alter the final amino acid chain.

20.6.2 Frameshift

20.6.3 Missense Mutation

Missense mutations are types of point mutations where a single nucleotide is changed to cause substitution of a different amino acid. This in turn can render the resulting protein nonfunctional. Such mutations are responsible for diseases such as Epidermolysis bullosa.

20.6.4 Nonsense Mutation

20.7 Further reading

20.7.1 Books

- Jones, S. 1993. *The Language of the Genes.* Harper Collins ISBN 0006552439.

20.7.2 Websites

- Wikipedia: Mutation[4]
- `http://www.evowiki.org/Mutation`

20.8 Original notes

- "Rare" change in nucleotide sequence
  - Somatic vs germline
- only those in germline are heritable
  - Point mutation
- Single nucleotide change
  - Change in gene position

4 `http://en.wikipedia.org/wiki/Mutation`

- Transposition
- Chromosomal rearrangement
- Mutagenic agents
- Raw material for evolutionary change

20.9 Point mutation

- Ionizing radiation
  - UV light induces thymine dimers
- Reparable
- Error during DNA synthesis
- Movement of transposons
  - McClintock
- Chemical mutagens
- May alter
  - Properties of promoter, enhancer
  - Amino acid sequence of polypeptide

20.10 Acquisition of genetic variability

- Mutation
- Sex (fusion of genomes)
- Recombination
  - Crossing over
- Reciprocal (may result in gene conversion)
- Unequal (gives rise to gene families)
  - Independent segregation
- Transposition by transposons
- Conjugation in bacteria
  - One way transfer from donor to recipient

20.11 Eukaryote genome

- Thousands of transposons
- Millions of transposon derived elements
  - LINES, SINES
- Above may constitute largest portion of genome
- Pseudogenes
- Tandem clusters (rRNA genes; nucleolus)
- Multigene families
- Single-copy genes (one copy per 1n)

20.12 Barbara McClintock

- Discovered transposons in perhaps greatest and ultimately most important intellectual endeavors in genetics
  - Maize
  - Worked alone
- Transposons: likely responsible for considerable evolution in eukaryotic genomes
- Likely origin of viruses

This text is based on notes very generously donated by Paul Doerder, Ph.D., of the Cleveland State University.

21 Recombinant DNA technology

- Revolutionized modern biology
  - Ability to manipulate genes in vitro
- Hybrid genes, including combining genes of different species
- Detailed study of gene function
  - Determine nucleotide sequences of genes and their regulators (deduce amino acid sequences of proteins)
- Genome projects: complete nucleotide sequence of >40 genomes, including human
- Made possible by convergence of:
  - discovery of restriction enzymes
  - genetics of bacteria and their plasmids

21.1 Recombinant DNA technology

- Uses
  - Detailed study of gene function
- Homeostasis, response to stress
- Development (birth defects)
  - Evolution of genes informs on evolution of life
- Human betterment
  - Medicine
- Identification, treatment of genetic disorders
- Molecular medicine: from deduced amino acid sequences, design better drugs
  - Foods
- Improve crop yield, resistance to disease
- Improve nutritional value
  - Forensics
- DNA fingerprinting: guilt or innocence

21.2 Restriction endonucleases

Originally found in bacteria to prevent invasion of viral DNA, cuts double stranded DNA that is unmethylated, will not cut newly synthesized DNA since hemi-methylated, a product of semi-conservative replication of DNA

- sever phosphodiester bonds of both polynucleotide strands in order to combine foreign DNA
  - create restriction fragments (restriction digestion)
  - 5' phosphate and 3' –OH at ends

- usually nucleotide specific target sequence
  - 4-6 bp most common, the more bases, then the more specific for recombination
  - cuts in or near sequence
  - ends
- sticky=overhanging ends, 5' or 3'
- blunt ends - straight cut, will anneal with any other blunt end in the presence of high ligase
- Hundreds of know restriction endonucleases, usually named after the bacteria that it was found in
  - e.g. EcoR1, Alu1, BAM, HIND3

21.3 Restriction endonucleases

Gene cloning

- Cloning:
  - Restriction digestion of DNA
  - insertion of restriction fragment into cloning vector
- Bacterial plasmid
- Bacterial virus
- Yeast artificial chromosomes
- Transformation of bacteria with recombinant plasmid, virus
- Screening for clone of interest by using reporter genes or resistance upon exposure to anti-biotic

21.4 Uses of cloned gene

- Determine nucleotide sequence and deduce amino acid sequence from genetic code
  - Submit to GenBank (available on WWW)
- Manipulate gene to study function
  - In vitro
  - In vivo
- Transgenic (recombinant) organisms
- Knockout organisms
- Medical and commercial uses

21.5 Other molecular procedures

- Polymerase chain reaction (Mullis)
  - Amplifies target DNA without cloning
  - Target amount can be single molecule
  - Amplified DNA can be sequenced, cloned, etc.
- Southern blotting
  - Used to identify restriction fragments carrying particular gene

  - Also used for DNA fingerprinting and RFLP analysis
- cDNA construction
  - Reverse transcription from mRNA template

21.6 RFLP(restriction fragment length polymorphism) analysis

- Basis of DNA fingerprinting using SNP - single nucleotide polymorphisms and repeats of DNA sequence
- Many uses
  - Criminal cases using multiple probes
  - Parentage
  - Species identification
  - Gene evolution
  - Species evolution

21.7 Sanger DNA sequencing

- Uses dideoxynucleotides (ddNTP), a template strand, DNA polymerase 1 (Also known as Kornberg enzymes) and dNTPs
  - Missing 3'-OH for nulceopjilic attack for elongation
  - DNA synthesis stops after one is incorporated into DNA fragment
  - ratio of ddNTP to dNTP determines likelihood of termination
- Manual method with 32P-labeled ddATP and 4 test tubes - ddATP, ddCTP, ddGTP, ddTTP
- Automated method using ddNTPs labeled with fluorescent dyes in capillary tube
  - Often done commercially

21.8 Automated sequencing

Typical machine

- - 2 hour sequencing run
  - 600-1000 bases per sample
  - multiple samples
- Up to 500,000 bases per day (12 hr)
- Data processed by computer
- In big labs, sequencing reactions also are automated

21.9 Genome projects

- Determine entire nucleotide sequence of genome

- >40 genomes sequenced
  - Helicobacter pylori
  - Escherichia coli
  - Saccharomyces cerevisiae
  - Caenorhabditis elegans
  - Drosophila melanogaster
  - Homo sapiens (first rough draft)
- Computer identifies all genes, based on properties of genes (e.g., start/stop codons, introns, etc.).

21.10 Biochips

- Microarray of DNA fragments, size of postage stamp; can be expensive, but has decreased in cost

Microarray chips contain wells of DNA that code for specific genes that uses the concept of hybridization with the gene of interest to see if a gene is expressed or is present.

- Designed to detect:
  - mutated genes (SNPs)
  - expressed genes
- Instant DNA profile ("GATTACA")

21.11 DNA chip controversies

- Medicine
  - Risks and informed consent for gene replacement therapy
  - Alteration of human gene pool
  - Parental choice
  - Privacy
- Genetically modified foods
  - Safety
  - Labeling
- Forensics
  - Mandatory tests
  - Reliability standards

21.12 Gene patenting

- Techniques to study and manipulate genes are patented (e.g., cloning and PCR)
- Should genes be patented?
  - Are they the intellectual property of the discoverer?
  - Don't they belong to all of us?
  - Should indigenous peoples be compensated for useful genes extracted from their local plants and fungi?

21.13 Stem cells

- Totipotent cells from early embryo
  - grow into any tissue or cell type
- Recombinant genes can be introduced
- Considerable use in analyzing gene expression in mice
- Possible therapeutic use in humans
- Very controversial

This text is based on notes very generously donated by Paul Doerder, Ph.D., of the Cleveland State University.

22 Classification of Living Things

22.0.1 Classification of Living Things & Naming

With so many flora and fauna on planet Earth, there must be a method to classify each organism to distinguish it from others so it can be correctly identified. Classification does not only apply to biology. For example, supermarkets and grocery stores organise their products by classifying them. Beverages may occupy one aisle, while cleaning supplies may occupy another. In science, the practice of classifying organisms is called **taxonomy** (Taxis means arrangement and nomos means law). The modern taxonomic system was developed by the Swedish botanist Carolus (Carl) Linneaeus (1707-1788). He used simple physical characteristics of organisms to identify and differentiate between different species.

Linneaeus developed a hierarchy of groups for taxonomy. To distinguish different levels of similarity, each classifying group, called **taxon** (pl. taxa) is subdivided into other groups. To remember the order, it is helpful to use a mnemonic device. The taxa in hierarchical order:

- Domain - Archea, Eubacteria, Eukaryote
- Kingdom - Plants, Animals, Fungi, Protists, Eubacteria (Monera), Archaebacteria
- Phylum
- Class
- Order
- Family
- Genus
- Species

The domain is the broadest category, while species is the most specific category available. The taxon Domain was only introduced in 1990 by Carl Woese, as scientists reorganise things based on new discoveries and information. For example, the European Hare would be classified as follows:

Eukaryote --> Animal --> Chordata --> Mammalia --> Lagomorpha --> Leporidae --> Lepus --> *Lepus europaeus.*

Binomial nomenclature is used to name an organism, where the first word beginning with a capital is the genus of the organism and the second word beginning with lower-case letter is the species of the organism. The name must be in italics and in Latin, which was the major language of arts and sciences in the 18th century. The scientific name can be also abbreviated, where the genus is shortened to only its first letter followed by a period. In our example, *Lepus europaeus* would become *L. europaeus'.*

Taxonomy and binomial nomenclature are both specific methods of classifying an organism. They help to eliminate problems, such as mistaken identity and false assumptions, caused by common names. An example of the former is the fact that a North American robin is

quite different from the English robin. An example of the latter is the comparison between crayfish and catfish, where one might believe that they both are fish when in fact, they are quite different.

Nomenclature is concerned with the assignment of names to taxonomic groups in agreement with published rules.

22.0.2 Eukaryotes & Prokaryotes

Recall that there are two basic types of cells: **eukaryotes** and **prokaryotes**.

Eukaryotes are more complex in structure, with nuclei and membrane-bound organelles. Some characteristics of eukaryotes are:

- Large (100 - 1000 μm)
- DNA in nucleus, bounded by membrane
- Genome consists of several chromosomes.
- Sexual reproduction common, by mitosis and meiosis
- Mitochondria and other organelles present
- Most forms are multicellular
- Aerobic

Prokaryotes refer to the smallest and simplest type of cells, without a true nucleus and no membrane-bound organelles. Bacteria fall under this category. Some characteristics:

- Small (1-10 μm)
- DNA circular, unbounded
- Genome consists of single chromosome.
- Asexual reproduction common, not by mitosis or meiosis.
- No general organelles
- Most forms are singular
- Anaerobic

22.0.3 The Three Domains

The three domains are organised based on the difference between eukaryotes and prokaryotes. Today's living prokaryotes are extremely diverse and different from eukaryotes. This fact has been proven by molecular biological studies (e.g. of RNA structure) with modern technology. The three domains are as follows:

Archea (Archeabacteria) consists of archeabacteria, bacteria which live in extreme environments. The kingdom Archaea belongs to this domain.

Eubacteria consists of more typical bacteria found in everyday life. The kingdom Eubacteria belongs to this domain.

Eukaryote encompasses most of the world's visible living things. The kingdoms Protista, Fungi, Plantae, and Animalia fall under this category.

22.0.4 The Six Kingdoms

Under the three domains are six kingdoms in taxonomy. The first two, **Plants** and **Animals**, are commonly understood and will not be expounded here.

Protista, the third kingdom, was introduced by the German biologist Ernst Haeckel in 1866 to classify micro-organisms which are neither animals nor plants. Since protists are quite irregular, this kingdom is the least understood and the genetic similarities between organisms in this kingdom are largely unknown. For example, some protists can exhibit properties of both animals and plants.

Fungi are organisms which obtain food by absorbing materials in their bodies. Mushrooms and moulds belong in this kingdom. Originally, they were part of the plant kingdom but were recategorised when they were discovered not to photosynthesise.

Eubacteria are bacteria, made up of small cells, which differ in appearance from the organisms in the above kingdoms. They lack a nucleus and cell organelles. They have cell walls made of peptidoglycan.

Archae (or Archaebacteria) are bacteria which live in extreme environments, such as salt lakes or hot, acidic springs. These bacteria are in their own category as detailed studies have shown that they have unique properties and features (ex. unusual lipids that are not found in any other organism)which differ them from other bacteria and which allow them to live where they live. Their cell walls lack peptidoglycan.

22.0.5 Origins of Diversity

The diversity in our planet is attributed to diversity within a species. As the world changed in climate and in geography as time passed, the characteristics of species diverged so much that new species were formed. This process, by which new species evolve, was first described by British naturalist Charles Darwin as **natural selection**.

For an organism to change, genetic mutations must occur. At times, genetic mutations are accidental, as in the case of prokaryotes when they undergo asexual reproduction. For most eukaryotes, genetic mutations occur through sexual reproduction, where meiosis produces haploid gametes from the original parent cells. The fusion of these haploid gametes into a diploid zygote results in genetic variation in each generation. Over time, with enough arrangement of genes and traits, new species are produced. Sexual reproduction creates an immense potential of genetic variety.

One goal of taxonomy is to determine the evolutionary history of organisms. This can be achieved by comparing species living today with species in the past. The comparison in anatomy and structure is based on data from development, physical anatomy, biochemistry, DNA, behaviour, and ecological preferences. The following are examples of how such data is used:

- Anatomy:

Although a horse and a human may look different, there is evidence that their arm structures are quite similar. Their arms' sizes and proportions may be different, but the anatomical structures are quite similar. Such evidence reveals that animals in different taxa may not

be that different. Biological features from a common evolutionary origin are known as **homologous**.

- Development
- Biochemistry:

Biochemical analysis of animals similar in appearance have yielded surprising results. For example, although guinea pigs were once considered to be rodents, like mice, biochemistry led them to be in their taxon of their own.

22.0.6 Phylogeny, Cladistics & Cladogram

Modern taxonomy is based on many hypotheses' of the evolutionary history of organisms, known as **phylogeny**. As with the Scientific Method, scientists develop a hypothesis on the history of an animal and utilise modern science and technology to prove the phylogeny.

Cladistics is a classification system which is based on phylogeny. Expanding on phylogeny, cladistics is based on the assumption that each group of related species has one common ancestor and would therefore retain some ancestral characteristics. Moreover, as these related species evolve and diverge from their common ancestor, they would develop unique characteristics. Such characteristics are known as **derived characteristics**

The principles of phylogeny and cladistics can be expressed visually as a **cladogram**, a branching diagram which acts as a family (phylogenetic) tree for similar species. A cladogram can also be used to test alternative hypotheses for an animal's phylogeny. In order to determine the most likely cladogram, the derived characteristics of similar species are matched and analysed.

22.0.7 Classification of Living Things Practice Questions

1. If taxonomists had to select an existing kingdom to reclassify, which of the six would most likely be chosen? Why?

2. Complete the following without consulting external sources:

a) The species *caudatum* is in the family *Paramecidae*. What would be the binomial name of this organism?

b) Give the abbreviation of the binomial name.

3.

a) Irish moss belongs to the genus *Chondrus*. The name for this species is *crispus*. Give the binomial name.

b) Give the abbreviation of the binomial name.

4. Humans and chimpanzees are alike. Which of the following data would most accurately prove this correct?

a) biochemistry

b) DNA

c) appearance

d) development

e) A, B, C

5. Which of the following is out of order?

a) Kingdom --> Phyllum --> Class

b) Class --> Family --> Order

c) Family --> Order --> Genus

d) Genus --> Species

e) A, C

f) A, B, D

g) B, C

6. A taxonomist discovers Organism A and Organism B and wishes to classify them. Which of the following choices is the most informative?

a) Both organisms are brown.

b) Both organisms have a tail.

c) Both organisms have ears.

d) Both organisms are nocturnal.

7. DNA analysis is usually done using DNA found in a cell's mitochondria, and not in a cell's nucleus. From your knowledge of mitosis, explain why this is so.

1. Arachbacteria 3.a) Chondrus crispus b) C. cripus 4. B 5. G 6. B

22.1 Introduction

Viruses are the smallest biological particle (the tiniest are only 20 nm in diameter). However, they are not biological organisms so they are not classified in any kingdom of living things. They do not have any organelles and cannot respire or perform metabolic functions. Viruses are merely strands of DNA or RNA surrounded by a protective protein coat called a **capsid**. Viruses only come to life when they have invaded a cell. Outside of a host cell, viruses are completely inert.

Since first being identified in 1935, viruses have been classified into more than 160 major groups. Viruses are classified based on their shape, replication properties, and the diseases that they cause. Furthermore, the shape of a virus is determined by the type and arrangement of proteins in its capsid. Viruses pathogenic to humans are currently classified into 21 groups.

Viruses can also attack bacteria and infect bacterial cells. Such viruses are called **bacteriophages**.

22.2 Viral Replication

As previously stated, viruses are not a biological life form so they cannot reproduce by themselves. They need to take over a functioning eukaryotic or prokaryotic cell to replicate its DNA or RNA and to make protein coat for new virus particles.

In order to enter a cell, a virus must attach to a specific receptor site on the plasma membrane of the host cell. The proteins on the surface of the virus act as keys which fit exactly into a matching glycoprotein on the host cell membrane. In some viruses, the attachment protein is not on the surface of the virus but is in the capsid or in the envelope.

There are two forms of viral replication: the **lytic cycle** and the **lysogenic cycle**.

22.2.1 Lytic Cycle

1. Attachment: The virus binds to specific receptors on the host cell.
2. Entry: There are two ways in which a virus can enter cells. Firstly, the virus can inject its nucleic acid into the host cell. Secondly, if a virus is contained in an envelope, the host cell can phagocytosise the entire virus particle into a vacuole. When the virus breaks out of the vacuole, it then releases its nucleic acid into the cell.
3. Replication: The virus's nucleic acid instructs the host cell to replicate the virus's DNA or RNA.
4. Assembly: New virus particles are assembled.
5. Lysis and Release: The virus directs the production of an enzyme which damages the host cell wall, causing the host cell to swell and burst. The newly formed virus particles are now released.

22.2.2 Lysogenic Cycle

1. Attachment: Similar to Lytic Cycle
2. Entry: Similar to Lytic Cycle
3. Incorporation: The viral nucleic acids is not replicated, but instead integrated by genetic combination (crossing over) into the host cell's chromosome. When integrated in a host cell this way, the viral nucleic acid as part of the host cell's chromosome is known as a **prophage**.
4. Host Cell Reproduction: The host cell reproduces normally. Subsequent cell divisions, daughter cells, contain original father cell's chromosome embedded with a prophage.
5. Cycle Induction: Certain factors now determine whether the daughter cell undergoes the lytic or lysogenic cycle. At any time, a cell undergoing the lysogenic cycle can switch to the lytic cycle.

The reproduction cycle of viruses with RNA and no DNA is slightly different. A notable example of a RNA-based virus is HIV, a retrovirus.

Retrovirus reproductive cycle

1. The retrovirus force RNA into cell, by either one of the two methods of entry (See above).
2. In the retrovirus are reverse transcriptase enzymes, which catalyses the synthesis of a DNA strand complementary to the viral RNA.
3. Reverse transcriptase catalyses a second DNA strand complementary to the first. With these two strands, the double-stranded DNA can be created.
4. DNA is then incorporated into the host cell's chromosomes. Similar to the concept of a prophage, this incorporated DNA is called a **provirus**. However, the provirus never leaves the host cell, unlike a prophage.
5. The infected host cell undergoes the lytic or lysogenic cycle.

22.3 Viral Genome

The genome of a virus consists of DNA or RNA, whose size and configuration vary. The entire genome can exist as a single nucleic acid molecule or several nucleic acid segments. Also, the DNA or RNA may be single-stranded or double-stranded, and either linear or circular.

Not all viruses can reproduce in a host cell by themselves. Since viruses are so small, the size of their genome is limiting. For example, some viruses have coded instructions for only making a few different proteins for the viruses' capsid. On the other hand, the human genome codes for over 30,000 different proteins. Therefore, the lack of coded instructions cause some viruses to need the presence of other viruses to help them reproduce themselves. Such viruses are called **replication defective**.

Lastly, it is worthy to note that 70% of all viruses are RNA viruses. As the process of RNA replication (with enzymes and other organelles of the host cell) is more prone to errors, RNA viruses have much higher mutation rates than do DNA viruses.

22.4 Viruses Practice Questions

1. As the name implies, the Tomato Spotted Wilt Virus targets tomatoes. Would it be possible for this virus to target other fruits as well? Explain.
2. If a DNA and a RNA virus both infected somatic cells, which virus would be more difficult to detect?
3. Many people have had cold sores, which are caused by infection with the herpes simplex virus. One characteristic of cold sores is that after a period of inactivity, they will reappear many times during the course of a person's life. Which cycle would the herpes simplex virus undergo?
4. Chicken pox is a common, non-fatal disease usually acquired in adolescence and caused by the varicella zoster virus. In adulthood, many people suffer from shingles, an altered form of the varicella zoster virus. Which cycle would the varicella zoster virus have undergone?
5. Would an antibiotic work for a person suffering from a cold of flu? Explain.

Answers to Viruses Practice Questions[1]

For Eubacteria, please visit General Biology/Classification of Living Things/Eubacteria[2].

22.5 Archaea

- Proposed as separate group from (eu)bacteria by Carl Woese
  - based on structure and metabolic pathways
  - inhabit extreme environments
  - unique branched lipids in membrane
- Share traits with both eukaryotes and eubacteria, e.g., RNA polymerase, introns
- Biochemically diverse
- Economically important
  - Taq polymerase used in PCR

22.5.1 Types

- Methanogens
- Halophiles
- Thermophiles

Underground bacteria

- Metabolism
  - built around inorganic energy sources
- e.g., basalt reacts with H_2O to release hydrogen which is catalytically combined with CO_2 to form carbohydrate (akin to photosynthesis)
- may result in deposit of minerals
- Unresolved problems
  - Did bacteria move downward from surface or did they first evolve there, protected from harsh surface conditions?
  - Could bacteria be ejected into space in rocks?

22.6 Prokaryote evolution

- Tentative, subject to change
- Derived largely from molecular systematics (rRNA sequences)
- Note: most bacteria can't be cultured, thus hard to study! (Studied by PCR of water/soil samples)

1 http://en.wikibooks.org/wiki/%2FAnswers

2 http://en.wikibooks.org/wiki/General%20Biology%2FClassification%20of%20Living%20Things%2FEubacteria

22.7 Domains of life: characteristics

This text is based on notes very generously donated by Paul Doerder, Ph.D., of the Cleveland State University.

22.8 Introduction

Out of the six kingdoms, Protista is the most diverse. This is the kingdom of organisms with strange, atypical characteristics. In essence, this kingdom is designated for organisms which do not belong in any other kingdom. The majority of protists are microscopic.

22.9 Classification of Protists

There are three phyla of protists, based on their type of nutrition.

1. **Protozoa** (animal-like protists) are heterotrophs that ingest or absorb their food.

2. **Algae** (plant-like protists) are autotrophs they get nutrition from photosythesis.

3. **Slime moulds** and **water moulds** (fungus-like protists) are also heterotrophs, like protozoa.

22.10 Protozoa

As heterotrophs, protozoa scavenge materials from their surroundings. Others are predators which actively hunt or ambush small organisms such as bacteria and other protozoa for a source of nutrition. Protozoa can be parasitic as well; they may live inside larger organisms, like humans. Most protozoa live as single cells, although a few form colonies.

Protozoa are generally difficult to identify due to their varied shape. They may appear as jelly-like blobs, spherical sunbursts, or a flattened leaf. Tiny blood parasites may be only 2 μm long. On the other hand, shell-covered marine may be 5 cm or more in diameter.

Furthermore, different protozoans have their own complex life cycles. The complexity has led certain organisms to be mistakenly classified for other species.

Nevertheless, protozoa can move, and so, they are classified based on their methods of locomotion.

Characteristics of Protozoa :

- About 30,000 species known
- About 10,000 species are pathogenic, including some of the worst human diseases
- heterotrophic
- highly variable in form and life cycle

- mostly unicellular
- range in size from 0.005 mm to 50 mm
- lack cell walls

they love environment and each other.....

22.11 Algae

Algae are much simpler than protozoa. They are aquatic and contain chlorophyll. Algae can exist as a single cell or as giant seaweeds 60 m in length. Formerly, algae were classified as plants but this was incorrect as algae lack parts of true plants: leaves, stems, roots, xylem, and phloem. Since algae belong in the kingdom Protista, algae is a broad term used to denote all aquatic eukaryotes which photosynthesise; algae can differ in size and shape as well.

There are six phyla of algae:**chlorophytes (green algae)**, **phaeophytes (brown algae)**, **rhodophytes (red algae)**, **chrysophytes (diatoms)**, **pyrrophytes (dinoflagellates)**, and **euglenophytes (euglenoids)**.

22.11.1 Chlorophytes

Chlorophytes resemble plants the most. Like plants, their cell walls contain cellulose and they store food in reserve as starch. Chlorophytes can be unicellular or multicellular. Most chlorophytes use flagellae for some locomotion.

22.11.2 Phaeophytes

Phaeophytes are nearly all multicellular marine organisms, which are known to us as seaweeds. They have cell walls composed of cellulose and alginic acid (a substance similar to pectin). The cellulose and alignic acid help to retain water and prevent seawood from drying out when exposed to air at low tide.

Since phaeophytes live in a tidal environment, they have large, flat fronds (a large leaf) which can withstand pounding by waves. Their bases strongly anchor the algae to the rocky seabed and prevent them from being washed out to sea. Phaeophytes are usually found in areas of cold water.

22.11.3 Rhodophytes

Rhodophytes are typically found in warmer seawater, and are more delicate and smaller than brown algae (phaeophytes). Rhodophytes are also able to grow at deeper depths in the ocean, since red algae absorb green, violet, and blue light, the wavelengths of which penetrate the deepest below the water surface. They also have mucilaginous material to resist drying.

22.11.4 Chryosophytes

Chryosophytes are the most abundant unicellular algae in the oceans. They are also one of the biggest components of plankton, a free-floating collection of microorganisms, eggs, and larvae. As photosynthetic organisms, they produce a significant amount of atmospheric oxygen.

The reproduction cycle of chryosophytes is particularly interesting. Note that diatoms reproduce both asexually and sexually. Since diatoms have a rigid cell wall with an outer layer of silica (found in sand and glass), the daughter cells produced by mitosis must fit inside the original cell wall. Therefore, each generation of diatoms is smaller than the one before. The reduction in size continues until the diatoms produce sexually, producing a zygote which eventually grows to the original size as it matures.

22.11.5 Pyrrophytes

Pyrrophytes are unicellular, photosynthetic, and mostly aquatic. They have protective coats composed of stiff cellulose. They are more easily identifiable, due to the presence of two flagellae. The longer flagellae propels the dinoflagellate, while the second shorter, flatter flagellae functions as a rudder.

Some species of pyrrophytes are **zooxanthellae**. Since they lack cellulose plates, they make their home in coral reefs and animals, such as sea anemones, and molluscs. In returning the favour of sheltering them, dinoflagellates provide carbohydrates to their host through photosynthesis. This is why there are nutrient-rich coral reefs in malnutritions water.

A negative aspect of pyrrophytes is that under certain conditions, species of dinoflagellates reproduce rapidly to form a **harmful algal bloom (HAB)**, known as a red tide if dinoflagellates are the cause. Such pyrrophytes can produce toxins which may injure or kill wildlife, and additionally any consumers of contaminated wildlife.

22.11.6 Euglenophytes

Like pyrrophytes, euglenophytes are small unicellular freshwater organisms with two flagella. They are mainly autotrophic or heterotrophic, depending if they have a red, light-sensitive structure called an **eyespot**.

22.12 Slime molds & Water molds

There are two phyla of slime moulds and one phylum of water moulds.

22.12.1 Oomycotes (Water moulds)

Oomycotes are filamentous organisms which resemble fungi, in that they live as saprotrophs. Oomycotes differ from other moulds with the presence of spores and their sexual life cycle.

22.12.2 Myxomycotes (Plasmodial slime moulds)

Myxomycoties are visible to the naked eye as tiny slug-like organisms which creep over decayed and dead matter. This streaming blob containing many nuclei is called a **plasmodium**.

22.12.3 Acrasiomycotes (Cellular slime moulds) and its reproductive cycle

Acrasiomycotes exist as individual amoeboid cells with one nucleus each. When in unfavourable conditions, each acrasiomycete cell gathers together to form a **pseudoplasmodium**.

Reproductive Cycle:

1. One acrasiomycete cell joins with others to form a pseudoplasmodium.

2. The pseudoplasmodium shrinks and forms a smaller plasmodium.

3. The plasmodium migrates to a suitable environment.

4. The plasmodium develops a sporangia, where original parental nuclei has divided by meiosis into haploid spores to be germinated.

5. When favourable conditions arise, the spores germinate and are carried away by animals or the wind.

6. Cycle repeats.

22.13 Protists Practice Questions

1. Which of the following adjectives describe the major food source of protozoa?

a) chemoautotrophic

b) photoheterotrophic

c) chemoheterotrophic

d) heterotrophic

e) A, C, D

f) C, D

2. The protozoan *Giardia lamblia* can inhabit a human body's intestinal tract and cause gastroenteritis.

a) Give the abbreviated binomial name of this protozoan.

b) Would the relationship between this protozoan and human being be mutualistic, commensalistic, or parasitic?

3. Found in many products, such as Petri dishes, agar is made from mucilagnious material in seaweed. Of the six phyla of algae, which phyllum/phyla would agar be made from?

4. Which of the following adjectives describe the major food source of Euglenophytes without an eyespot?

a) photoautotrophic

b) photoheterotrophic

c) chemoautotrophic

d) chemoheterotrophic

e) B or C

f) C or D

5. Can coral reefs exist in nutrient-poor areas? Explain.

23 Multicellular Photosynthetic Autotrophs

23.1 Plants

- Multicellular
- Cellulose cell walls
- Chlorophylls a and b
- Develop from embryophyte
- Alternation of generations
- Major food source for terrestrial life
- Atmospheric O2 and CO2 balance
- Coal deposits
- Intimate association with mycorrhizal fungi
- >250,000 species (~500,000?)
- Taxonomy
  - State of flux
- DNA sequencing
- Developmental studies
  - Division (old literature) = phylum (new literature)
  - ~12 phyla, 9 of which are vascular plants

23.2 Plant phyla

Phyla are 12 groupings

23.3 Plant evolution

- Evolved from green algae, likely related to charophytes
- Evidence
  - DNA sequences
  - homologous chloroplasts: chlorophyll b and beta-carotene; thylakoids in grana;
  - Cellulose in both groups; also peroxisomes
  - Mitosis and cytokinesis similar
  - Sperm ultrastructure

23.3.1 Terrestrial adaptations

- Stomata: pores in leaves for exchange of gases; prevent desiccation
- Secondary metabolites:
  - cuticle: waxy coating to prevent H2O loss
  - lignin: hardens wood
  - sporopollenin: resistant polymer; coats pollen
  - predator defenses
- Embryonic development
  - gametangia in early plants
  - spores; seeds
- Mycorrhizae
- Water/food conducting systems

23.4 Plant phylogeny

23.5 Plant life cycles

- Alternation of generations
- Sporophyte
  - diploid
  - produces spores in sporangia
- Gametophyte
  - develops from spore
  - haploid
  - produces gametes in gametangia
- Haplodiplontic life cycle

23.6 Moss life cycle

23.7 Vascular plants

- Most have roots
- Aerial shoot systems
- Vascular tissue
  - xylem: water, mineral transport
  - phloem: food transport
- Lignin
- Branched sporophyte is dominant stage
  - amplified production of spores
  - evolution of complex plant bodies
- Dominated Carboniferous (360 my)

23.8 Vascular plant life cycles

- Homosporous (single type of spore)
- Heterosporous (two types of spore)

23.9 Pterophyta (ferns)

- Non-seed plant
- Sporophyte conspicuous (vascular tissue)
- Rhizome: ground stem, roots
- Fronds: leaves
- Sori: clusters of sporangia
- Motile sperm require external water for fertilization
- Originated in Devonian, 350 my

Tree fern Fern life cycle

23.10 Non-seed plants, continued

- Lycophyta: club mosses
  - E.g., Lycopodium ("ground pine")
  - Many species became extinct 270 my, once dominant (coal formations)
  - Gametophyte non-photosynthetic, nourished by fungi
- Arthrophyta: horsetails
  - Equisitum
  - Some fossil forms (300 my) were tree-size (coal)
  - Photosynthetic stems, no leaves
  - Silica deposits in epidermal cells

23.11 Seed plants

- 1st appeared in Devonian, 360 my
- Seed develops from ovule, protects embryo
  - withstands drought
  - dispersal is enhanced
  - no immediate need for water for germination
- Heterosporous
  - male gametophyte: arise from microspores
  - female gametophyte: arise from megaspores in ovule in ovary
- Two groups
  - gymnosperms
  - angiosperms

```
plant
```

23.12 Sporophyte/gametophyte

23.13 Megasporangium (nucellus)

- Key to seed development
- Nucellus: solid, fleshy, surrounded by integuments derived from sporophyte (seed coat)
- Entire structure called ovule
- Flower may have many ovules

23.14 Pollen

- Develop from microspores, become male gametophyte
- Protected by sporopellenin
- In most plants, sperm lack flagella (loss)
- Many mechanisms to transport pollen
  - wind
  - insects, birds,

23.15 Gymnosperms

- "naked seed"
  - ovule not fully enclosed by sporophyte at time of pollination
- Conifers, cycads, gnetophytes, Ginkgo
- Small, inconspicuous plants to giants like sequoia
- Conifers: to carry cones fv
  - male cones, Female conesvv
  - evergreen

23.16 Pine life cycle

23.17 Other Coniferophyta

- Cycadophyta: cycads
  - tropical, subtropical
  - flagellated sperm
- Gnetophyta
  - e.g., Ephedra, Mormon Tea
- Ginkgophyta: Ginkgo

- only one surviving species
- diocious (separate % and &trees)

23.18 Other gymnosperms

23.19 Angiosperms

- Flowering plants, Anthophyta
  - monocots- single seed leaf (grasses, lilies, etc.)
  - dicots- two seed leaves (roses, pulses, maples)
- More specialized xylem (water transport)
  - vessel elements
  - fiber cells
- Fossils date to 130 my
- Animal (e.g., insect) coevolution

23.19.1 Monocots vs dicots

23.20 Earliest angiosperm

- What is earliest angiosperm?
- Recent analysis of nucleotide and amino acid sequences suggests that Amborella, a tropical plant found only on the island of New Caledonia, is closest relative to flowering plants

23.21 Angiosperm flower

Insert non-formatted text hereInsert non-formatted text here

23.22 Angiosperm life cycle

This text is based on notes very generously donated by Paul Doerder, Ph.D., of the Cleveland State University.

23.23 Introduction

Although you may not recognise fungi, they are just as prevalent as plants and animals. Their spores are in the air which we breathe, fungi allow us to make bread, and mushrooms (a type of fungi) are eaten by us. A few types of fungi are unicellular. For example, yeasts live as individual oval or cylindrical cells. However, the majority of fungi live are multicellular. Their bodies are composed of **hyphae**, a network of fine filaments. In a mushroom, the

hyphae are densely packed so it is difficult to see the individual structures when a mushroom is eaten. However, a mushroom is only a specialised reproductive part of the whole fungus. The main part of the fungi is underground in a whole web of hyphae, called a **mycelium**.

In the mycelium, each fungal cell is separated from each other by a septum. Each fungal cell may have one or more nuclei and remains connected to the mycelium because the septa are porous, allowing cytoplasm to flow through the hyphae and fungal cell walls, made of a hard material called **chitin**. Some fungi do not have septa, and they appear to be large, branching, multinucleate cells.

23.24 Nutrition

Fungi are **saprophytes**. When they find a source of food (e.g. dead wood, orange peel) , they decompose it and digest it. The enzymes break down larger organic molecules in the substrate into smaller molecules. These smaller molecules diffuse into the fungus, where they are used to allow growth and repair.

Fungi which feed on living cells are parasitic. For example, athlete's foot grows on the human foot. These kinds of fungi produce hyphae called **haustoria**, which can penetrate host cells without immediately killing them.

However, they are friendlier species of fungi. Many fungi live symbiotically with plants or animals. For example, most trees have fungi living in close contact with their roots. In this relationship, known as a **mycorrhiza**, there are many benefits:

- Growing around the plant roots and often entering plant cells, the hyphae absorb minerals from the soil and release them in the roots. The fungi gets its source of food (organic nutrients) while delivering food to the plant.
- The mycelium here would increase the surface area, thus the absorptive surface, of the plant roots.
- The fungal cells help to maintain air and water flow in the soil around the roots.
- The fungi may prevent other potentially pathogenic fungi to attack the tree.

23.25 Fungal Reproduction

Fungi can reproduce in two ways. Firstly, they make asexually produce through **fragmentation**. This occurs when pieces of hyphae are broken off, which then grow into new mycelia.

The second method is by spores. Spores are lightweight structures and windblown designed to be transported over long distances and by many mediums, such as on the bodies of insects and birds. They are additionally light enough to be blown away for hundreds of kilometers. Spores may be asexual and sexual. Their sexual properties can be analysed to classify the four phylla of fungi.

23.26 Types of Fungi

23.26.1 Zygospore Fungi (Zygomycetes)

This phyllum includes bread moulds and other saprotrophs. Comparable to bacteria, this phyllum prefers asexual reproduction over sexual reproduction.

1. Two haploid hyphae of opposite types, also known as **mating strain +** and **mating strain -**, combine and fuse together.

2. **Plasmogamy**, the union of the two parent hyphae, occurs and results in the creation of a heterokaryotic (n + n) **zygosporangium or zygospore**. Note that the zygospore is NOT diploid yet; the haploid nuclei are simply clumped together.

3. Immediately, a thick wall develops around the zygospore to protect it from drying and other hazards. The zygospore becomes dormant.

4. When conditions are favourable, the zygospore absorbs water and undergoes **karyogamy** (n + n = 2n), where the haploid nuclei contributed by the two parents fuse to produce diploid zygosporangia.

5. The now diploid zygosporangium then undergoes meiosis to form haploid sporangia.

6. Through asexual reproduction of fungi (See above for more information), the spores from the sporangia germinate and grow into new mycelia.

7. Back to step #1.

23.26.2 Club Fungi (Basidiomycetes)

This phyllum increases mushrooms and shelf fungi. In many ways, the reproduction stages of this phyllum is similar to that of zygomycetes.

1. Two haploid hyphae of opposite types, also known as **mating strain +** and **mating strain -**, combine and fuse together.

2. Plasmogamy takes place, and a dikaryotic mycelium forms. The dikaryotic mycelium grows faster then the haploid parental mycelia.

3. Environmental factors cause the dikaryotic mycelium to form compact masses which develop into **basidiocarps**, short-lived reproductive structures. An example is the mushroom.

4. The basidiocarp gills are lined with terminal dikaryotic cells called **basidia**, which then undergo karyogamy.

5. The basidia are now diploid. They undergo meiosis to develop haploid **basidiospores**, a term referring to a basidiomycete's spores.

6. Still remaining on the basidiocarp, the haploid basidiospores eject, fall from the basidiocarp, and are dispersed by the wind when mature.

7. In a favourable environment, the basidiospores germinate and grow into short-lived haploid mycelia.

8. Back to Step #1.

23.27 Key Terms

synapomorphy

23.28 Introduction

What makes an animal an animal?

If animals are a monophyletic taxon, then animals should be able to be defined by synapomorphies, (shared, derived characteristics). Ideally, we would NOT define this or any taxon using symplesiomorphies (shared ancestral or primitive characteristics) or homoplastic characters (the independent evolution of similarity, or "convergent evolution"). See pages 654 - 656 and Fig. 32.6 in your text to review these concepts. As you consider the characteristics listed below, ask yourself whether or not each is a synapomorphy.

23.29 Characteristics of an Animal

- There is no one universally accepted definition of an animal. The following treatment follows your text, beginning on page 876.
  - Animals:
    - Are multicellular, heterotrophic eukaryotes ...
    - Lack the distinctive cell walls of plants & fungi
    - Share unique characteristics ...
    - Share certain reproductive characteristics ...
    - Other commonly used definitions ...

Animals are multicellular heterotrophic eukaryotes

- Unfortunately, none of these traits is exclusive to animals:
  - Plants, fungi, and some algae are multicellular.
  - Many bacteria, protists, and all fungi are heterotrophic.
  - Everything other than bacteria and archaea are eukaryotic.
- Moreover, all three of these characteristics also apply to fungi.
  - However, there is a difference here between animals and fungi. Animals generally take in their food through ingestion, or eating and swallowing something. Fungi are absorptive heterotrophs; they secrete their digestive enzymes onto their food, and then absorb the resulting nutrients.

Animals share unique characteristics

- Only animals have muscle tissue and nervous tissue.
- Only animals have collagen, a structural protein
- Only animals have the following types of intercellular junctions: (See pages 135 - 139, Figure 7.15 in your text for more information on these junctions.)

- Tight junctions (sealing function)
- Desmosomes (anchoring function)
- Gap junctions (communication function)

Animals share certain reproductive characteristics

- Most animals reproduce sexually, with the diploid stage dominating.
- In most animals, a small, haploid, flagellated, motile sperm fertilizes a larger, haploid, nonmotile egg to form a diploid zygote.
- Mitotic division of the zygote yields a blastula stage, followed by a gastrula stage. A synapomorphy? This feature could be another "unique characteristic" shared by animals.
- Development may be direct to adult form, or there may be a sexually immature stage (or stages) that are morphologically & ecologically distinct from the adult called a larva (plural: larvae).

Other commonly used definitions or characterizations

- It is surprisingly difficult to find two texts that agree on a precise definition of an animal. Here are a few perspectives from some other texts.
  - Animals are multicellular eukaryotes; they are chemosynthetic heterotrophs that ingest their food.
  - Animals are motile, though many are secondarily sessile. Gametes usually are produced in multicellular sex organs, and the zygote passes through embryonic stages that include a blastula.
  - Animals are organisms that are multicellular, with more than one type of cell. They are heterotrophic. They reproduce sexually (at least sometimes), with a zygote formed from two different haploid gametes. They go through a developmental stage called a blastula.
  - Animals are not photosynthetic, have no cell wall, and no hyphae or mycelia. (What would a cladist think of this definition of the taxon Animalia?)

What kinds of animals are there?

- Kingdom Animalia generally is recognized to have approximately 30 phyla ...
  - There is relatively little dispute over the number of phyla recognized; however, the phylogenetic relationships among the phyla are hotly debated.
  - Molecular techniques for assess similarity based on nucleotide sequences in nucleic acids are providing valuable new perspectives on this question.
- Remember that two animals in different phyla generally are considered to be more different from each other than are animals within one phylum (e.g., nematodes are more different from annelids than humans are from sharks).

This text is based on notes very generously donated by Ralph Gibson, Ph.D.[1] of the Cleveland State University[2].

1 http://bgesweb.artscipub.csuohio.edu/faculty/gibson.htm
2 http://www.csuohio.edu

23.30 Introduction to animal phyla

There currently are almost 40 recognized phyla.

Phylum — Number of Species — Common Name

- Placozoa[3] — 1
- Monoblastozoa[4] — 1
- Rhombozoa[5] — 50
- Orthonectida[6] — 50
- ***Porifera***[7] — 9,000 — sponges (figures)
- ***Cnidaria***[8] — 9,000 — corals (figures)
- Ctenophora[9] — 100 — comb jellies
- ***Platyhelminthes***[10] — 20,000 — flatworms (figures)
- Nemertea[11] — 900 — ribbon worms (figures)
- ***Rotifera***[12] — 1,800 — rotifers (figures)
- Gastrotricha[13] — 450 — gastrotrichs
- Kinorhyncha — 150 — kinorhynchids
- ***Nematoda*** — 12,000 — roundworms (figures)
- Nematomorpha — 230 — horsehair worms
- Priapula — 15
- Acanthocephala — 700 — (figures)
- Entoprocta — 150
- Gnathostomulida — 80
- Loricifera — 35
- ***Annelida*** — 15,000 — segmented worms (figures)
- Sipuncula — 250 — peanut worms (figures)
- Echiura — 135
- Pogonophora — 145 — beard worms
- Vestimentifera — 8 — beard worms
- ***Arthropoda*** — 957,000 — arthropods (figures)
- Onychophora — 80
- Tardigrada — 400 — water bears
- Pentastomida — 95 — tongue worms
- ***Mollusca*** — 100,000 — molluscs (figures)
- Phoronida — 15
- Ectoprocta — 4,500 — sessile zooids
- Brachiopoda — 335 — lampshells

3 http://en.wikibooks.org/wiki/Placozoa
4 http://en.wikibooks.org/wiki/Monoblastozoa
5 http://en.wikibooks.org/wiki/Rhombozoa
6 http://en.wikibooks.org/wiki/Orthonectida
7 http://en.wikibooks.org/wiki/Porifera
8 http://en.wikibooks.org/wiki/Cnidaria
9 http://en.wikibooks.org/wiki/Ctenophora
10 http://en.wikibooks.org/wiki/Platyhelminthes
11 http://en.wikibooks.org/wiki/Nemertea
12 http://en.wikibooks.org/wiki/Rotifera
13 http://en.wikibooks.org/wiki/Gastrotricha

- ***Echinodermata*** — 7000 — echinoderms (figures)
- Chaetognatha — 100 — arrow worms (figures)
- Hemichordata — 85 — acorn worms
- ***Chordata*** — 50,000 — chordates (figures)

23.31 Phylum Porifera

Figure 6 Sponges

Name means "pore-bearing".

This phylum consists of the sponges. The number of species is estimated to be between 5,000 and 10,000. All are aquatic and almost all are marine.

Animals in this phyla have no true tissues, which means, for example, that they have no nervous system or sense organs. Although sponges are multicellular, they are described as being essentially at a cellular level of organization. They are sessile as adults, but have a free swimming larva.

Their bodies are porous. They are filter feeders; water flows in through many small openings, and out through fewer, large openings. They have inner and outer cell layers, and a variable middle layer. The middle layer often is gelatinous with spiny skeletal elements (called spicules) of silica or calcium carbonate, and fibres made of spongin (a form of collagen).

Choanocytes are flagellated cells lining the inside of the body that generate a current, and trap and phagocytize food particles.

Their cells remain totipotent, or developmentally flexible: they can become any type of cell at any point in the sponge's development. This allows for the great regenerative power sponges have.

Sponges are an ancient group, with fossils from the early Cambrian (ca. 540 mya) and possibly from the Precambrian. Sponges often are abundant in reef ecosystems. They somehow are protected from predators (spicules? bad taste?).

Many organisms are commensals of sponges, living inside them. Some sponges harbor endosymbiotic cyanobacteria or algae (dinoflagellates, a.k.a. "zooxanthellae").

23.32 Phylum Cnidaria

See text pages 886 - 889.

Name comes from the Greek knide- meaning "nettle".

This phylum They have one opening, which serves as both mouth and anus. The body wall has an outer ectoderm, an inner endoderm, and a variable undifferentiated middle layer called mesoglea or mesenchyme that may be jelly-like. The mesoglea is **NOT** considered to be true mesoderm and so the Cnidaria are described as diploblastic. Tentacles usually extend from the body wall around the mouth/anus.

Figure 7 Jellyfish Development

There are two basic body plans: the polyp[14] and the medusa[15]. The polyp is sessile and attaches to substrate by the aboral end (i.e., the end away from the mouth). The medusa ("jellyfish") is a floating form, and looks like an upside-down version of the polyp. Some cnidarians only have the polyp stage, some have only the medusa stage, and others have both.

14 http://en.wikipedia.org/wiki/polyp
15 http://en.wikipedia.org/wiki/Medusa%20%28biology%29

The typical life cycle of a cnidarian involves what is called "alternation of generations": an alternation between an asexual polyp stage and a sexual medusa stage.

The tentacles are armed with cnidae (or nematocysts), small intracellular "harpoons" that function in defense and prey capture. When fired, the cnidae deliver a powerful toxin that in some cases is dangerous to humans. The phylum is named after the cnidae.

Cnidarians have no head, no centralized nervous system, and no specialized organs for gas exchange, excretion, or circulation. They do have a "nerve net."

Many cnidarians have intracellular algae living within them in a mutualistic symbiotic relationship (Dinoflagellates = zooxanthellae). This combination is responsible for much of the primary productivity of coral reefs.

There are three main classes in the phylum

- Class Hydrozoa[16] (hydras and Portugese man-of-war are well-known but atypical examples of this Class)
- Class Scyphozoa[17] (jellyfish)
  - The medusa stage is dominant and the polyp stage often is reduced.
- Class Anthozoa[18] (sea anemones, most corals)
  - No medusa (jellyfish) stage, so sexual reproduction occurs in the polyp stage in this group. The polyps also can reproduce asexually, which is how individual "corals" grow.

23.33 Phylum Platyhelminthes

See text pages 890 - 893.

Name means "flat worm"

Most members of this phylum are parasitic (flukes and tapeworms), but some are free living (e.g., planaria). There are about 20,000 species.

They are dorsoventrally compressed (i.e., "flat").

Animals in this phylum are acoelomate, triploblastic, bilaterally symmetrical, and unsegmented. Platyhelminths have a simple anterior "brain" and a simple ladder-like nervous system. Their gut has only one opening. Flatworms have NO circulatory or gas exchange systems. They do have simple excretory/osmoregulatory structures (protonephridia or "flame cells").

Platyhelminths are hermaphroditic, and the parasitic species often have VERY complex reproductive (life) cycles.

There are four main classes of platyhelminths:

- Class Turbellaria[19] (mostly free living flatworms, e.g., planaria)

16 http://en.wikipedia.org/wiki/Hydrozoa
17 http://en.wikipedia.org/wiki/Scyphozoa
18 http://en.wikipedia.org/wiki/Anthozoa
19 http://en.wikipedia.org/wiki/Turbellaria

- Class Monogenea[20] (parasitic flukes)
- Class Trematoda[21] (parasitic flukes, e.g., liver fluke and the human blood fluke, Schistosoma)
- Class Cestoda[22] (tapeworms)
  - Cestodes are endoparasitic in the gut of vertebrates. They do not have a mouth or digestive system.

23.34 Phylum Rotifera

See text page 900

The Rotifers. The name means "wheel bearing," a reference to the corona, a feeding structure (see below).

They are triploblastic, bilaterally symmetrical, and unsegmented. They are considered pseudocoelomates.

Most less than 2 mm, some as large as 2 - 3 mm.

Rotifers have a three part body: head, trunk foot. The head has a ciliary organ called the corona that, when beating, looks like wheels turning, hence the name of the phylum. The corona is a feeding structure that surrounds the animal's jaws. The gut is complete (i.e., mouth & anus), and regionally specialized. They have protonephridia but no specialized circulatory or gas-exchange structures.

Most live in fresh water, a very few are marine or live in damp terrestrial habitats. They typically are very abundant. There are about 2,000 species.

Parthenogenesis, where females produce more females from unfertilized but diploid eggs, is common. Males may be absent (as in bdelloid rotifers) or reduced. When males are present, sexual and asexual life cycles alternate. Males develop from unfertilized haploid eggs and are haploid. Males produce sperm by mitosis which can fertilize haploid eggs, yielding a diploid zygote that develops into a diploid female. Sexual reproduction occurs primarily when living conditions are unfavorable.

Most structures in rotifers are syncytial ("a mulitnucleate mass of protoplasm not divided into separate cells," or "a multinucleated cell") and show eutely (here, "constant or near-constant number of nuclei").

23.35 Phylum Nematoda

See text pages 894 - 895.

Name from the Greek for "thread".

20 http://en.wikipedia.org/wiki/Monogenea
21 http://en.wikipedia.org/wiki/Trematoda
22 http://en.wikipedia.org/wiki/Cestoda

This phylum consists of the round worms. There are about 12,000 named species but the true number probably is 10 - 100 times this!

These animals are triploblastic, bilaterally symmetrical, unsegmented pseudocoelomates. They are vermiform, or wormlike.

In cross-section, they are round, and covered by a layered cuticle (remember this cuticle !!). Probably due to this cuticle, juveniles in this phylum grow by molting. The gut is complete. They have a unique excretory system but they lack special circulatory or gas-exchange structures. The body has only longitudinal muscle fibers. The sexes are separate.

Nematodes can be incredibly common, widespread, and of great medical and economic importance. They are parasites of humans and our crops. They can live pretty much anywhere.

> In one rotting apple, there can be up to 90,000 nematodes, and in one tablespoon of coastal mud, there can be 236 species of nematodes!

Nematodes can be free living or important parasites of our crops, or of humans and other animals. They have become very important in development studies, especially the species Caenorhabditis elegans, presumably due to its small size and constancy of cell number (eutely - 959 cells in C. elegans).

23.36 Phylum Annelida

See text pages 906 - 909.

Name means "ringed", from the Greek annulatus.

This phylum consists of earthworms, leeches, and various marine worms given many different names (e.g., sand worms, tube worms). There are about 12,000 - 15,000 species.

Animals in this phylum are triploblastic, bilaterally symmetrical, segmented coelomates. Development is typically protostomous. They have a complete circulatory system, and a well-developed nervous system. Typically, each segment has paired epidermal "bristles" (setae or chaetae).

Most are marine but they are successful occupants of almost anywhere sufficient water is available. They can be free living, parasitic, mutualistic, or commensalistic.

Major advances of this phylum include the true coelom, segmentation, both longitundinal and circular muscles, a closed circulatory system and, for most, a more advanced excretory system (metanephridia).

There are three main classes of Annelids

- Class Oligochaeta (earthworms)
- Class Polychaeta (marine worms)
- Class Hirudinea (leeches)

23.37 Phylum Arthropoda

Figure 8 Arthropods

Name means "jointed feet".

This phylum consists of spiders, ticks, mites, insects, lobsters, crabs, and shrimp, and is the largest of all the phyla. So far, over 1 million species have been named, and it is likely that the true number out there is 10 - 100 times greater.

This phylum also includes the extinct trilobites, which were prevalent in the Paleozoic era. Because of their exoskeletons, these animals fossilized well and over 4000 species have been named.

These animals are triploblastic, bilaterally symmetrical, segmented, protostome coelomates. The coelom is generally reduced to portions of the reproductive and excretory systems. They have an open circulatory system.

The most notable advancement of this phylum is a rigid exoskeleton. It has major implications in these organisms' locomotion, flexibility, circulatory systems, gas exchange systems, and growth. It also was partially responsible for the ability of the arthropods to move on to land.

There are several major groupings of arthropods:

Major subgroups include:

- The chelicerates (eurytperids, horseshoe crabs, scorpions, spiders, ticks) have clawlike feeding appendages. They lack antennae and usually have simple eyes.
- The Trilobites...they get their own grouping

- The uniramians (centipedes, millipedes, insects) have one pair of antennae and unbranched (uniramous) appendages.
- The crustaceans (crabs, shrimp, lobsters, barnacles and many others) have two pairs of antennae and branched (biramous) appendages.

Major Classes Include

- Class Arachnida (mites, scorpions, spiders, ticks)
- Class Diplopoda (millipedes)
- Class Chilopoda (centipedes)
- Class Insecta (insects)
- Class Crustacea (crabs, crayfish, lobsters, shrimp)

23.38 Phylum Mollusca

See text pages 900 - 905.

Name means "soft".

This phylum consists of snails, slugs, bivalves, chitons, squids, octopus, and many others. About 110,000 species

All molluscs have a similar body plan:

- A muscular foot, usually used for movement.
- A visceral mass, containing most of the internal organs.
- A mantle, a fold of tissue that drapes over the visceral mass and secretes the shell, if present.
- Most have a radula, or a rasping organ to scrape food.

Molluscs are bilaterally symmetrical, or secondarily asymmetrical. They are coelomates, but the coelom generally has been greatly reduced; the main body cavity is a hemocoel. Development is typically protostomous. The gut is complete with marked regional specialization. Large, complex, metanephridia (excretion).

Many molluscan life cycles include a trochophore larva. This stage also is characteristic of annelids.

There are several major classes of molluscs:

- Class Polyplacophora (chitons)
- Class Gastropoda (snails, slugs, nudibranchs)
- Class Bivalvia (clams, mussels, scallops, oysters)
- Class Cephalopoda (squids, octopuses, chambered nautiluses)

23.39 Phylum Echinodermata

Name means "spiny skin"

This phylum consists of sea stars, brittle stars, sea urchins, and sea cucumbers.

Echinoderms are mostly sessile or very slow moving animals.

As adults, they are radially symmetrical, but in the larval stage, they are bilaterally symmetrical. They are considered deuterostomes.

Echinoderms are unique in that they have a water vascular system composed of a system of fluid-filled canals. These canals branch into the tube feet, which function in feeding, locomotion, and gas exchange.

There are six major classes of echinoderms:

- Class Asteroidea (sea stars)
- Class Ophiuroidea (brittle stars)
- Class Echinoidea (sea urchins, sand dollars)
- Class Crinoidea (sea lilies)
- Class Holothuroidea (sea cucumbers)

23.40 Phylum Chordata

Name means "the chordates", i.e., these animals have a notochord at some stage in their lifecycle.

This phylum consists of tunicates, lancelets, and the vertebrates.

There are four major features that characterize the phylum Chordata.

- A notochord, or a longitudinal, flexible rod between the digestive tube and the nerve cord. In most vertebrates, it is replaced developmentally by the vertebral column. This is the structure for which the phylum is named.
- A dorsal hollow nerve cord which develops from a plate of ectoderm that rolls into a tube located dorsal to the notochord. Other animal phyla have solid nerve cords ventrally located. A chordate nerve cord splits into the central nervous system: the brain and spinal cord.
- Pharyngeal slits, which allow water that enters through the mouth to exit without continuing through the entire digestive tract. In many of the invertebrate chordates, these function as suspension feeding devices; in vertebrates, they have been modified for gas exchange, jaw support, hearing, and other functions.
- A muscular, postanal tail which extends posterior to the anus. The digestive tract of most nonchordates extends the length of the body. In chordates, the tail has skeletal elements and musculature, and can provide most of the propulsion in aquatic species.

Chordates have a segmented body plan, at least in development. This segmentation evolved independently from the segmentation of annelids.

Three subphyla make up the phylum Chordata:

- Subphylum Urochordata (tunicates): the adults are enclosed in a tunic made of a carbohydrate much like cellulose. They squirt water out of an excurrent siphon. Urochordates are characterized by errant (mobile and active) larvae and sessile adults. All are filter feeders. The only "chordate" characteristics retained in adult life are the pharyngeal slits.

Larval urochordates look more like adult cephlochordates & adult vertebrates than adult urochordates.

- Subphylum Cephalochordata: Cephalochordates are known as lancelets because of their blade-like shape; they are also known as amphioxus. They are marine animals and usually live on the bottom, but can swim.
- Subphylum Vertebrata (vertebrates) ...

Formally, the phyla Urochordata and Cephalochordata are considered invertebrates.

23.40.1 Subphylum Vertebrata

Vertebrata refers to the presence of vertebrae and a vertebral column.

This subphylum includes most of the animals with which most people are familiar.

Vertebrates show extreme cephalization.

The notochord generally is replaced by the cranium & vertebral column in adults.

Neural Crest Cells

Later in development, these give rise to many cells of the body, including some cartilage cells, pigment cells, neurons & glial cells of the peripheral nervous systems, much of the cranium, and some of the cells of the endocrine system.

Some scientists would like to classify the neural crest as the fourth germ layer.

Neural crest cells come from the dorsal edge of the neural plate, thus ectoderm.

24 Chordates

The phylum Chordata includes three subphyla. These include vertebrates and invertebrate chordates.

24.1 Characteristics

Notochord: the rod-shaped supporting axis found in the dorsal part of the embryos of all chordates, including vertebrates

Flexible, non-collapsible rod dorsal to the gut/coelom and below the nervous system, hydrostatic, fluid wrapped in tough connective tissue. As bone does not compact, muscles tensed on one side result in movement instead of shortening the animal. This allows much better locomotion than do cilia for larger animals in water, a crucial victory for later success.

Pharyngeal slits: Slits in the pharynx originally used to gather food, water enters the mouth, passes through pharynx and out gill-like slits, passing through a cavity called an antrium and then outside. In humans, present only in embryo.

Dorsal nerve cord: A neural tube dorsal to the notochord

Postanal tail: Elongation of the body and notochord, nerve cord and muscles past anus into tail, early locomotive function led to success.

Non-synapomorphic characteristics (not limited to chordates):

- bisymmetrical (bilateral symmetry)
- segmented muscles and bones

24.2 Subphylum Urochordata

The tunicates are located in this subphylum. Along with the subphylum Cephalochordata, these two subphyla make up the invertebrate chordates. Only the tunicate larvae have notochords, nerve cords, and postanal tails. Most adult tunicates are sessile, filter-feeders which retain their pharyngeal slits. Adult tunicates also develop a sac, called a tunic, which gives tunicates their name. Cilia beating within the turnicate cause water to enter the incurrent siphon. The water enters the body, passes through the pharyngeal slits, and leaves the body through the excurrent siphon. Undigested food is removed through the anus. Tunicates are hemaphrodites and can reproduce asexually through budding.In urochordates notochord is confined to larval tail.These lack cranium. These have an open type of circulatory system.Excretion is by neural gland,nephrocytes.

24.3 Subphylum Cephalochordata

The lancelets are located in this subphylum. Along with the subphylum Urochordata, these two subphyla make up the invertebrate chordates. Lancelets receive their name from their bladelike shape. They resemble fish but they are actually scaleless chordates only a few centimeters long. They spend most of their time buried in the sand with their mouths protruding. Fossils of lancelets have been found to be over 550 million years old.

Dropped out sessile stage, what was the larval stage is now sexually reproductive. Includes Branchiostoma ("amphioxus").

24.4 Subphylum Vertebrata

(Vertebra from Latin vertere, to turn). Characterized by separate bones or cartilage blocks firmly joined as a backbone. The backbone supports and protects a dorsal nerve cord. Vertebrates have tissues which are organized into organs which in turn are organized into organ systems.

All vertebrates share the following characteristics: - segmentation - a true coelom - bilateral symmetry - cephalization - a backbone - a bony skull - a closed circulatory system - chambered heart - two pairs of jointed appendages - tissues organized into organs

Vertebrate Organ Systems: - Nervous System - Circulatory System - Digestive System - Respiratory System - Reproductive System - Excretory System

- Vertebral column: Not present in higher vertebrate adults. (In humans, the gel-like, spongy core of the vertebral column is the only remainder. Ruptured or herniated disc is an injury to this.)
- Cranium: Composite structure of bone/cartilage. Two functions: 1. Supports sensory organs of head and 2. Encloses or partially encloses the brain.

What evolutionary relationship could we imagine between sessile echinoderms and the higher chordate animals?

Paedomorphic (child-form) hypothesis: basically, evolution of sexual reproduction in what had previously been a larval life stage, or the retention of at least one juvenile characteristic into the adult (adult = sexually reproducing) stage. Some scientists believe that this occurred in a proto-chordate animal lineage. Maybe chordates (and vertebrates) arose from sessile (attached) ancestors. Selection in these proto-chordates maybe began to favor more time in the larval stage, as feeding was more successful or mortality lower in this stage. As larvae got bigger physics shows that the cilia become less efficient for locomotion, favoring the undulating motion allowed by a notochord.

Is this hypothesis crazy? A similar example of this today is Epemeroptera, the mayfly, which has almost abandoned its adult stage. Its one-year lifespan is mostly larval with just a brief day of reproduce-and-die as an adult, which doesn't even have usable mouthparts.

Tunicate (sea squirt) larva has all four chordate characteristics, although adult sessile ("attached").

24.4.1 Class Agnatha

"jawless fish"

- Ostracoderms: extinct Agnathans which had primitive fins and massive plates of bony tissue on their body.
- Cyclostomes: "circle mouth" - a group of Agnathans which is still alive in the form of lampreys and hagfish.

Appeared approximately 500 million years ago and dominated the oceans for about 100 million years. The first group of fish to appear. They had neither jaws, paired fins, nor scales, but they were the first organisms with a backbone.

Class Acanthodia

"spiny fish" Appeared about 430 million years ago. An extinct class of fish that developed jaws with bony edges. They had internal skeletons made of cartilage and some bone.

Class Placodermi

Appeared about 410 million years ago, dominated the sea for about 50 million years. An extinct class of fish with massive heads.

24.4.2 Class Chondrichthyes

"cartilaginous fish" Appeared about 400 million years ago with bony fish. Includes sharks, skates and rays, and chimaeras. Their skeletons are made of cartilage strengthened by the mineral calcium carbonate.

The main characteristics and distinguishing features of this class: - gills - single-loop blood circulation - vertebral column - presence of placoid scales on their bodies - internal skeleton of cartilage - paired, fleshy pectoral and pelvic fins - asymmetrical tail fin prevents sinking - fatty liver provides neutral buoyancy - visceral clefts present as separate and distinct gills - no external ear - oviparous - internal fertilization - ectoderms - cold blooded

24.4.3 Class Osteichthyes

"bony fish" Appeared about 400 million years ago with cartilaginous fish. Includes about 95% of today's fish species.

Subclass Sarcopterygii

fleshy-finned fishes. Fins have bones and muscles, homologous to our limbs.

Order Dipnoi

lung fishes, two groups isolated when continents separated

Order Crossopterygii
includes coelacanths and rhipodistians, gave rise to amphibians, had lungs which evolved into a swim bladder in bony fishes, and labyrinthodont teeth, characterized by complex folding of enamel.

- Skeleton made of bone, jaws, fins, most with scales, two-chambered heart.

24.4.4 Class Amphibia

means "both lives", aquatic larvae, terrestrial adult Amphibians: - Legs - Lungs - Double-Loop Circulation - Partially Divided Heart - Cutaneous Respiration (Breathes through Skin)

Order Salientia
frogs (jumping) (aka Anura)

Order Urodela
salamanders (tailed)

Labyrinthodont amphibians: oldest known amphibians, inherited characteristic teeth from crossopterygii ancestor, had stocky, aquatic larvae.

Amphibians have limbs instead of fins. Girdles and vertebral column now more substantial and connected, support body on legs.

Lisamphybia: no scales, "smooth", eggs with no shell, laid in water (water-reliant).

Amphibians gave rise to cotylosaurs, from which arose dinosaurs, turtles, lizards, and therapsids.

Class Reptilia

amniotic egg allowed freedom from water, shelled egg. (Amnion: protection). Reptiles have four extra-embryonic membranes:

- Amnion: supports aquatic environment inside egg in fluid sac.
- Allantois: allows gas exchange and elimination of wastes.
- Chorion: gas exchange
- Yolk sac: only one of the four left over from amphibian ancestor

Reptiles cold-blooded, or ectothermic, meaning that their heat come from their environment. Sometimes defined as all amniotes that are not birds or mammals.

Reptiles can be classified by skull structure into four groups:

- Anapsid
- Synapsid
- Diapsid
- Euryasid

Refers to number of holes in the skull. Cotylosaurs had Anapsid skull

Dermatocranium: from bony outer skull structure, precursor to human cranium.

Subclass Anapsidia

Subclass Testudinata
turtles[1], terrapins

Subclass Diapsida
dinosaurs[2], snakes[3], most stuff

Subclass Synapsida

Order Therapsids

Subclass Diapsida
includes Ichthyosaurs, marine reptiles convergent on dolphins; Plesiosaurs, ancient sea monsters; Squamates, including lizards and snakes; and Thecodonts, which gave rise to

- birds
- dinosaurs
- crocodilians

Dinosaurs: broken into two groups, based on hip structure

- Saurischia: lizard hips (gave rise to birds [!]), ancestrally bipedal
- Ornithischian: bird hips, ancestrally quadripedal

Crocodilians: come from archosaurs, the only extant (still living today) archosaur descendant. Ancestrally bipedal, secondarily quadripedal.

Synapsids: refers to joined (Greek syn-, together with) parts of skull. Led eventually to mammals. Synapsid pelycosaur >> therapsid >> mammals

Pelycosaur: Sail-backed dinosaur, legs not spread out like lizard but more pillar-like and under body, allowing greater activity and competence in motion, pendulum like rather than constant push-up. Teeth differentiated into different types, for pre-processing of food needed by higher metabolism. Skull changes, bone histology, suggestions of warm-bloodedness.

1 http://en.wikibooks.org/wiki/turtle
2 http://en.wikibooks.org/wiki/dinsosaur
3 http://en.wikibooks.org/wiki/snake

Class Aves

arose late Jurassic, early Cretaceous. Feathers, skeleton modified for flight. Feathers: epidermal derivative, made of keratin (like fingernails). Carpometacarpis: bears primary flight feathers, parallel to hand parts. Keeled sternum: breastbone, powerful one needed to support flight muscles. Strong, light, occasionally hollow bones. All birds lay eggs (as contrasted to reptiles, which have developed live birthing over 100 independent times.) Why are there no live-bearing birds? Early birds had teeth, lost them. With mammals, only exothermic animals.

Archaeopteryx: "ancient wing", Jurassic bird-reptile, very dinosaur-like. Good fossils found in Zolenhoffen, German sandstone mine with fine sand, shows feathers clearly, found shortly after Darwin's publication and used to support his hypothesis. Thick, heavy bones and no sternum, bony tail, not a good flyer but did have primary flight feathers.

Archaeornithes: includes archaeopteryx.

Paleognathae: gave rise to Australian flightless birds.

Neognathae: remaining live birds.

Class Mammalia

Two unique characteristics, or synapomorphies:

- Hair
- Mammary glands

(don't fossilize well)

Three skeletal characteristics (fossilize)

- Lower jaw only one bone, the dentary (several in reptiles)
- Three bones in middle ear: malleus, incus, stapes (reptiles have one or two, never three)
- Joint between upper and lower jaws between dentary and squamosal of skull (in reptiles this joint is between other bones)

Mammals basically have a synapsid skull design inherited from ancestor

Non diagnostic characteristics (not unique to mammals):

- Warm-blooded
- Skin glands: sweat glands and oil-producing sebaceous glands
- Large nasal cavities (because of high metabolism) Clean, warm and humidify air
- Heterodonty (differentiated teeth)
- Diphiodonty: two sets of teeth: baby and adult ("deciduous" teeth, drop out) (reptile teeth are continually replaced)

Subclass Protheria

monotremes (Greek mon-, one; and trema, hole), or egg-laying mammals, have one opening for excretion and urination.

Subclass Theria
Metatheria: Marsupials (opossum, kangaroo. . .) Eutheria: Placental mammals (all common mammals)

Marsupium: (from Greek marsypion, purse or pouch). Gestation period much shorter than in Eutherian mammals, but after leaving the uterus the tiny offspring crawls into a pouch where it completes development latched onto a teat.

Recent molecular (read: genetic) evidence suggests that two different mammal groups may have developed live-bearing ability separately. Instead of being a "rough draft" for placental-style live bearing, perhaps the marsupial pouch approach is another solution to the same problem. Advantage: in tough times the parent can pitch out the offspring and increase its own chance of survival.

25 Tissues and Systems

26 Epithelial tissue

Comes from various sources, **ectodermal** or **endodermal** material. Cell sheet lines a surface or body cavity. One side, called **freesurface or Apical**, is exposed to

- animal **interior** (forming the **lumen**) or
- **exterior** of its **body cavity**.

The other side rests on the **basal layer**.

Epithelial tissue is not penetrated by blood vessels.

Two categories:

- **sheets**
- **glands**

Classified on two features:

- **simple**, (a single layer of cells),
- **stratified**, (more than one cell layer.)

Cell shape at free surface:

- **squamous** (broad and flat)
- **cuboidal** (spherish)
- **columnar** (tall and thin)

Simple squamous epithelium

usually lines body cavities and vessels,alveoli, glomeruli of kidney; in blood and lymph vessels called endothelium; in body cavities called mesothelium (serosae): parietal serous membranes line body wall, visceral serous membranes cover organ

Simple cuboidal epithelium

in ducts like kidney and salivary glands.

Simple columnar epithelium

nonciliated type lines digestive tract, ciliated type lines some regions of uterine tubes and lungs

Stratified squamous epithelium

(important) lines mouth, esophagus,and vagina. Cells sometimes dead, flat and keratinized, making them resistant to abrasion. Stratified squamous epithelium changes to columnar squamous epithelium progressively down esophagus to the stomach.

Epidermis

from epithelium. Below this is **dermis**, thicker and with blood vessels.

Two specialized epithelia:

- **pseudostratified**
- **transitional**

Pseudostratified epithelia

lines the trachea (where it is ciliated)and the male urethra (where it is non ciliated), looks stratified but not.

Transitional epithelia

found only in bladder and urinary system. As it stretches it appears to go from 6 to 3 cell layers deep.

Glandular epithelia

(**gland**: group of cells that excretes something.. mostly derived from **epithelium**. Glands are classified into **endocrine** and **exocrine** by where they excrete.

Endocrine glands

secrete hormones into the **blood** without use of ducts.

Exocrine glands

secrete onto the **body surface** or into a **cavity**, thru a **duct**. Exocrine substances include sweat, mucous, oil, and saliva. An exocrine gland is the **liver**, which secretes bile.

27 Connective tissue

This is a "grab bag" category of diverse tissue types. Functions include binding and supporting. Types include **bone**, **cartilage**, **fibrous connective tissue**, **blood** and **adipose** (fat) tissue.

If you took away everything in the body except the connective tissue, you'd still be able to see the basic form of the body.

Form: distinctive cells surrounded by a **cell matrix** made of **extra-cellular fiber** grounded in a **ground substance** (excluding blood)

Types:

1. **connectile connective tissues** (can be 1. **loose** or 2. **dense**)

2. **special connective tissue** (includes blood, bones and cartilage).

Fibroblasts form **connective tissue proper;**

chondoroblasts form **cartilage;**

osteoblasts form **bone;**

and **blood** is formed from various sources.

Ground substance: "unstructured" material that fills space between cells and contains fibers. Made of

1. **interstitial fluid** (bathes cells)

2. **proteoglycans** (protein core with attached polysaccharides, glycoaminoglycans or GAGs such as chondroitin sulfate, keratin sulfate, and hyalronic acid, whose consistency is syrupy to gelatin-like)

3. **cell-adhesion proteins** (connect connective tissue cells to the fibers).

Fibers of **connective tissue**:

1. **Collagen** (flexible protein resistant to stretching, tensile strength, most abundant protein in animals, white)

2. **elastin** (rubbery, resilient protein, in dermis, lungs, blood vessels, yellow when fresh)

3. **andreticulin** (like collagen).

Loose connective tissue: found beneath skin, anchors muscles,nerves etc. Include **fibroblasts**, **macrophages**, **mast cells**,and **adipose cells**. **Fibers** include **collagen** and **elastic fibers**. Ground substance is "syrupy". Adipose included.

Dense connective tissue: largely densely packed fibers of **collagen** or **elastin** regularly or irregularly arranged. Forms **tendons** and **ligaments**, coverings of muscles, capsules around organs and joints, and **dermis** of skin.

Cartilage vs. bone

Feature	**Bone**	**Cartilage**
cell type	osteocytes	chondrocytes
ground substance	calcium phosphate	chondroitin sulfate
vascularization	vascular	avascular
micro architecture	highly ordered	less organized
units called	osteons	
fibrous sheath	peristeum	perichondrium

Cartilage: There are three cartilage types:

1. **hyaline** cartilage

2. **fibrocartilage** (fibrous cartilage)

3. **elastic** cartilage

Hyaline cartilage: most widespread cartilage type, in adults forms articular surfaces of long bones, rib tips, rings of trachea, and parts of skull. Mostly **collagen**, name refers to **glassy** appearance. In embryo, bones form first as **hyaline** cartilage, later **ossifies**. Found in tracheal rings. Few collagen fibers.

Fibrous cartilage: have lots of **collagen** fibers. Found in intervertebral discs, pubic symphesis. Grades into dense **tendon** and **ligament** tissue.

Elastic cartilage: springy and elastic. Found in internal support of external ear and in epiglottis, yellow when fresh.

Chondrocites (cartilage cells) rely on diffusion for nutrients, as cartilage has no direct blood supply, and no enervation (nerves). Can be loaded with calcium salts.

Bone: Specialized connective tissue, calcium phosphate arranged in highly ordered unit called **osteon**, or **Hyvercian system**. Concentric rings around central canal with blood vessels and enervation (nerves). Bone varied, not all vertebrate bone is even cellular. Our concern: simple pattern for mammals.

Lacuna (spaces in which osteocytes found); **canaliculi** (little canals) bigger diagonal cells, layers of bone called **lamellae.**

Three types of bone cells, ending in

-**blast**, (mend bone)

-**cyte** (fortify bone)

-**clast** (tear down bone)

Classified by

1. **appearance** (**spongy** vs. **hard**)

2. **where found** (outside or inside)

3. **how it is formed** (endochondral cartilage model forms first and then is ossified, and entramembranous, bone forms directly without cartilage precursor)

Example of **endochondral bone formation**: long bone begins to ossify from center shaft, calcified region expands and cuts off diffusion of nutrients as bone replaces cartilage. In young mammals, secondary ossification centers then form at bone ends, growth has stopped by sexual maturity as all primary bone is ossified. In other animals, bones continue growing throughout their lifetime.

Three types of **intramembrous bone**:

1. **dermal** bone

2. **sessamoid** bone

3. **perichondral** bone.

Dermal bone forms skull, shoulder/pectoral girdle, and integument, descended from dermal armor of ancestor. Comes from **mesoderm**, in dermis of skin.

Sessamoid bones: form directly in tendons. Example: kneecap, also in wrist. Deals with stress.

Perichondral bone means **"around cartilage,"** forms around cartilage or bone. Functions in bone repair and in ossification of endochondral bone.

Bone remodeling and repair: bone has mineral structure, and develops tiny fractures, which, under stress, can lead to larger fractures. To combat this, bone is constantly replaced. **Osteoclasts** channel through existing bone, tear down and leave behind **osteoblasts** and **lacuna**, leaving **osteocytes**. Continually resets mineral structure of bone, and is preventative maintenance.

When bone broken, callus forms in open ends, **periosteum** gives rise to new bone with calcium and new bone matrix, leaves irregular mend. Later, osteoblasts continue fixing over time and slowly removing imperfection.

28 Muscle tissue

Mesodermal in origin, muscle has several functions: **supply force for movement**, **restrain movement**, proper **posture**, act on viscera (internal organs) for **peristalsis** (moving food down digestive tract), give body **shape**, form **sphincters**, (such as in esophagus, between stomach and intestine, large and small intestine, in anus), in **sheets** of muscles, affect ai**r flow** in and out of lungs, line **blood vessels** and play vital role in **circulation**.

Secondary roles: **heat production** (shivering a specialized heat production to supplement metabolism).

Muscles co-opted to other non-original functions: sharks detect electrical field created by fish muscles. Some fish formed electric organs, create current strong enough to repel predators or stun prey. Other fish can use field as "radar" to see things and communicate with other animals. (Evolved independently in different groups).

Different classifications: by **color**, (red or white) **location**, nature of **nervous system contro**l (voluntary or involuntary), **embryonic origin**, or by **general microscopic appearance** (striated, smooth, and cardiac.)

Striated muscle (or **skeletal** muscle): under **voluntary control**. Individual cells called **fibers**, grouped into **fascicle**. **Myofibrils** founding one cell made of even smaller **myofilaments**. Each striated cell very long and **multi-nucleated**. Fibers joined end to end to form longer **composite fibers**. **Sarcomeres**: repeating units make up **myofibrils**. Two kinds of **myofilaments**, thick kind made up of **myosin** and thin of **actin**. Striations visible in light microscope, smaller part only with electron microscope.

Cardiac muscle: occurs only in **heart**. Light banding visible under light microscope. Each band short, principally mononucleate (occasionally dinucleate) often branched, joined together with intercollated discs. Involuntary. Waves of contraction spread through intercollated discs. Initiated by nerve stimulation or can originate in the heart itself (useful in heart transplants.)

Smooth muscle: no striations visible with light microscope. Almost entirely **visceral** function: digestion, sphincters, urogenital tracts, piloerectory muscles (make hairs stand up), lungs. Non-voluntary control. Slow and sustained action. Each cell mononucleate, short, fusiform (spindly) in shape, cells usually uniform in size.

Striated muscle contraction: Muscle broken into units called fascicles, in units of myofibrils. Repeating units called sarcomeres, consisting of two kinds of myofilaments:

1. thick, **myosin** filament

2. thin, **actin** filament.

Sarcomere: Thick and thin filaments interspersed in ordered grid.

Sliding filament theory: thick and thin filaments move past each other in opposite direction, shortening length. Longer muscles contract more rapidly than short ones (see cell bio for details).

Myosin molecule: two polypeptides twisted together with two globular heads at end.

Myosin filament: many slender myosin molecules together.

Actin filament: chain of actin single, **tropomyosin** strands with repeated **globular troponin**, and with **actin**. All play role in muscle contraction. Myocin heads have sites that bind to actin. Actin filaments have many regular sites that can bind to myosin.

Troponin has **four site**s:

1. one to bind **myosin**

2. one for **actin**

3. one for **tropomyocin**

4. one for **calcium ions**

Nerve signal reaches muscle, triggers release of chemical signal called neurotransmitter, that diffuses across cell membrane (**sarcolimic reticulum**) and binds to receptors in it. Receptor is **acetylcholine**, ACH. When there is enough nerve signal, the message travels through t-line to sarcoplasmic reticulum to release calcium ions.

Lacking calcium, **tropomyosin** site blocked. In calcium, myosin binding sites exposed and heads bind to **actin** molecules, delivering force to move **fibers** in relation to each other. Myocin head then interacts with **ATP** to get "recocked", if myosin still exposed then it fires again and results in further muscle contration. If there is no further nerve signal, **sarcoplasmic reticulum** sequesters Ca+ ions again and no recocking occurs.

Quirari (or curare): known from movies, used in South America, blocks **acetylcholine receptors** in cell and causes skeletal paralysis. Victim dies of asphyxiation because he can't breathe.

Duchenne's muscular dystrophy: degeneration of **sarcolema**, plasma membrane of muscle cell unable to release signal and quickly atrophies.

Fast and **slow twitch fibers**: vertebrate muscle fiber. Terms relative within one group of animals. Differences related to differences in **enervation**, type of **myocin**, and **actin activation**.

Two parts of force generated by muscle: 1. **active** component 2. **elastic** component (energy stored in muscle when stretched by gravity or another force. Stored in muscle elastic tissue around tendons. Especially important in **limb oscillation**, like running, or trunk twisting, like fish swimming. Up to 90% of stored elastic energy can be recovered.)

How does a muscle match its power to its job? Two ways:

1. **rate modulation**, derived from frequency of nervous stimulation of muscle, force increases as frequency of stimulation increases up to point of tetanus.

2. **selective involvement of motor units**, a given neuron enervates a fixed number of muscle cells, (a motor unit), and force is increased by recruiting more motor units. Motor units may be small, such as in eye, or larger, like in leg muscle.

How do muscles grow stronger?

1. add more **myofilaments**, increases cross sectional area by up to 50%, more little ratchets working

2. proliferation in **blood vessels** and **connective tissue** around muscle

Muscle strength is relative to **cross sectional area**, not length. Not always feasible to add more cross sectional area.

Pinnate fibers: oriented obliquely (Y-shaped) to minimize muscle mass, in certain circumstances, like calf muscle. Spreads muscle out.

Velocity of shortening greater in long muscle than short. Why? Contraction tied to relation between fibers, and to total length of muscle. Both long and short muscles reach same percentage of contraction in same unit time, but distance covered by the longer muscle is greater.

Synergist muscles: muscles work together to produce motion in same general direction. Bicep shares work with brachialis.

Antagonist muscles: muscles that oppose each other. Bicep pulls forearm in, triceps pulls it back out.

Origin vs. insertion: origin is the end of the muscle that more fixed in its attachment to the body. The more movable end called insertion.

Fixators: muscles that act to stabilize a joint or lever system. Like upper arm when you clench your fist hard.

Flexors and extensors: applied mainly to limbs. Flexor bends one part relative to another about limb, extensor straightens it.

Adductor and abductor: adductor draws a limb toward the ventral surface. Abductor moves limb away from ventral surface. (Adduct: drawn toward; abduct: carry away).

29 Vertebrate digestive system

Functions to break down food into molecules small enough to absorb, or pass across digestive membrane.

Digestive tract: tube extending from lips of mouth to anus or cloacae in bird, reptile or monotreme.

Lumanal glands: empty into inner body cavity (lumen: inner surface).

Tract divided into three main regions: 1. buccal cavity 2. pharynx 3. alimentary canal

Alimentary canal divided into four regions: 1. esophagus 2. stomach 3. small intestine 4. large intestine

Accessory digestive glands, outside digestive tract proper, secrete into lumen of tract through ducts. Includes the **salivary glands**, **liver** and **pancreas**.

Buccal cavity, which includes **palate** and **tongue**, develops from infolding of **stomadeum**, or second opening of blastula, whereas the rest of the digestive tract develops from the **primitive** gut**.**

Teeth: capture and hold prey. In mammals in particular further process and break down food into small particles, increasing surface area available for enzymatic action.

Tooth anatomy: 1. **crown** projects above gum, 2. **root** below gum, 3. **enamel** is outer coating of crown, hardest surface in body, of epideral origin 4. **dentin**, below enamel, bone-like and forms bulk of tooth, is harder than bone and contains nerves and blood vessels. (Remember that mammals are **heterodontic**, with different types of teeth).

Pharynx: air passage for adult, gill slits in embryo. Important in lower vertebrates, site of gills. Features derived from **pharyngeal** region: first pharyngeal pouch gives rise to parts of the ear, other pouches give rise to various other structures.

Alimentary canal: epithelium lines lumen, glands secrete into lumen, longitudinal and circular muscles help digestive movements (**peristalsis**).

Esophagus: tube carries food from mouth to stomach. Expands to fit large **bolus** (lump of chewed food). Secretes mucus for lubrication. Birds have **crop** for storage, enlargement of esophagus.

Epiglottis: keeps food out of air tube, an evolutionary "kludge," or fix.

Stomach. Absorbs water, alcohol, nutrients, uses gastric juice with enzymes, mucous, HCl, released by **chief** and **parietal cells** (release protein enzymes) in gastric pits. **Rugae**: folds of stomach, disappear when full. **Sphincter** at both ends of stomach, control food passage.

Chyme: semi-digested food released to small intestine.

Small intestine: three regions, **duodenum**, **jejunum**, and **ileum**.

Duodenum site of most intestinal digestion. **Jejunum** and **ileum** do most of intestinal absorption. Ileum ends with another sphincter, **ileocolic valve** or ileosecal valve. Structure: Circular folds covered with **villi** (singular is villus).

Villi: finger-like cellular projections, covered with **microvilli**, tiny projections which increase surface area. Increases surface area by 900x, speeds **digestion** (break down) and **absorption** (taking in nutrients).

Large intestine: larger diameter, shorter length than small intestine. No villi. In mammals, forms large gentle loop, colon, empties into straight region, rectum, empties into outside world through anal sphincter. Colon: absorbs water left over, also absorbs vitamins released by bacteria which live there (vitamin K).

Food: made up of 1. **proteins**, 2. **fats**, 3. **carbohydrates** 4. **fibrous material**.

Digestive system breaks foods down. **Proteins** must be broken to **amino acids** to be absorbed. **Polysaccharides** to **monosaccharides**, **lipids** to **fatty acids** and **monoglycerides** to absorb.

Salivary glands in mouth, saliva contains mucous, salt and a few enzymes (amalase, begins starch breakdown). Snake venom from oral gland, mixture of toxins and digestive enzymes. Breaks down blood vessels and disables nervous system.

Stomach enzymes: released in inactive form, **zymogene**, converts to active form in lumen of gut. Transformation is triggered by another enzyme, or the stomach's low pH. **Pepsin** secreted as **pepsinogen** (-ogen means primitive form). Stomach glands secrete up to two or three liters a day of gastric juice, which is reabsorbed.

Chyme released to **duodenum**.

Small intestine has two major accessory glands: 1. **pancreas** 2. **liver**

Pancreas has **endocrine** and **exocrine** functions, releases large amounts of carbonate to neutralize acidic chyme, as intestinal enzymes work in neutral pH, and stuff to break down lipids and starch (zymogens, like tripsin)

Liver releases **bile**. Bile made from cholesterol, stored in **gall bladder**, released in duodenum, emulsifies fats.

Emulsify: keeps fats in tiny drops, which are suspended, increasing surface area and action of lipases. Protein and carbohydrates absorbed in intestine, taken to liver for processing. Fatty acids go to lymphatic system

Appendix: vestigial remnant. Much variation in digestive systems within mammals: herbivore, carnivore, insectivore, non-ruminant herbivore.

Rumen: four-chambered stomach of animals like cows (ruminant herbivores). Cellulose resistant to digestion, rely on microorganisms to break down cellulose. Some bacteria, protists and fungi can break down cellulose, almost no animals can. Bacteria break down cellulose in rumen, to be taken back to the mouth to chew their cud (ruminate). Later cow swallows to proceed with digestion. (Horses not like this).

Coprophagy: rabbits and other animals eat their own feces for the nutritious products of the cecum.

30 Circulatory system

Circulatory system functions

1. **Transportation**

```
 a. Respiration: gas exchange (O2 and CO2), overcomes limited rate
of
```

diffusion

```
 b. Nutrition:<br>
 c. Excretory: (remove metabolic wastes)<br>
```

2. **Regulation**

```
 a. Transport hormones<br>
 b. Regulate body temperature<br>
 c. Protection<br>
    i.  Blood clotting<br>
    ii.  Immune system (carries white blood cells)
```

Vasodilation: allows heat loss across epidermis, as seen in elephant ears, takes more blood to surface of body, sweating may accompany

Countercurrent heat exchange: used by dolphins in fins to conserve heat in cold water. Veins surround an artery, and blood returning to body absorbs heat from blood traveling out from body to fin, minimizing heat loss. Used by dogs in feet, etc.

Blood made of 1. **plasma** and 2. **formed cellular elements** (red and white blood cells, and platelets).

Plasma makes up 55% of blood volume. Cellular elements make up the other 45%.

Plasma makeup: 90% water, 7-8% soluble proteins (**albumin** maintains blood osmotic integrity, others clot, etc.) 1% electrolytes 1% elements in transit

Red blood cell (**erythrocyte**): contains **hemoglobin**, functions in oxygen transport. In mammals, red blood cells lose nuclei on maturation, and take on biconcave, dimpled, shape. No self repair, live 120 days. About 1000x more red blood cells than white blood cells. About 7-8 micrometers in diameter.

Hematocrit: proportion of blood volume that is occupied by cells, about 43% in humans on average. 48% for men and 38% for women.

White blood cells (**leukocytes**): Nucleated, about 10-14 micrometers in diameter, commonly amoeboid, escape circulatory system in capillary beds. Include basophils, eosinophils, neutrophils, monocytes, B- and T-cell lymphocytes.

Platelets (**thrombocytes**) Membrane bound cell fragments in mammals, no nucleus. In non-mammals, platelet role replaced by nucleated cells. Accumulate at site of broken blood vessels, form clots. Bud off special cells in bone marrow. 1-2 micrometers in diameter. 7-8 day life span, 1/10 or 1/20 as abundant as white blood cells.

Arteries: carry blood **away** from heart. Smallest tubes called **arterioles**, feed blood to capillaries.

Veins: return blood to heart. Smallest veins called **venules**.

Structure of arteries and veins, listed from inside (lumen) out: 1. **epithelium** (endothelium), 2. **elastic connective tissue fibers**, 3. **smooth muscle**, 4. **connective tissue**. Arteries have thicker elastic layer than do veins.

Capillaries, where exchange of materials occurs, are very thin and narrow, and red blood cells pass through single file. Capillaries are tiny but numerous, and their total volume is greater than that of supplying arteries.

Blood velocity drops in capillaries, picks back up in veins. Pressure highest in arteries, lower in capillaries and arteries.

Osmotic pressure draws interstitial fluid from blood in arterioles, but replaces it in venules.

One-way valves mean that blood can flow only one way, works with residual blood pressure and compression by skeletal muscles. Low pressure in thoracic cavity caused by breathing also helps move blood.

Lymphatic system: part of the immune system, a one-way, or **open**, system. Takes up **interstitial fluid** not taken up by venules.

Lymphatic structures:

1. lymphatic **capillaries**

2. lymphatic **vesicles**

3. **lymph nodes**

4. lymphatic **organs** (**spleen** and **thymus**)

Lymph: movement in mammals through one-way valves, similar to blood movement in veins. (Some non-mammals have lymphatic hearts of unknown embryonic origin. Frogs and salamanders have several.) Lymph rejoins cardiovascular system into a large vein near the heart via single large thoracic duct.

As lymph passes through system, passes **lymphocytes**, second part of immune system.

Heart: pumps blood, design varies between animals. In adult mammal,four chambers form two separate circulations

1. **pulmonary circulation** to and from lungs and

2. **systemic circulation** to and from tissues of body.

Everything in the heart comes in pairs: 2 atria, 2 ventricles (left and right).

Diagrams usually drawn as though animal were on its back.

Pattern of blood flow through heart: blood returning from **major veins** (vena cava) enters right **atrium**, contraction there delivers blood to right **ventricle** through a **tricuspid valve**, one of atrial ventricular valves (AV valve). Contraction of right ventricle drives blood through **semi lunar valve** into **pulmonary circuit** and to **lungs**.Blood return to heart in **pulmonary veins**, is oxygenated. Goes to left **atrium**, which contracts and delivers blood to left **ventricle**by way of **aortic semi-lunar valve**, then goes to **systemic circulation**.

Both **atria** and **ventricles** contract in unison, left is more powerful than right (to all system vs. just lungs).

Systole: heart contraction, **diastole**: heart relaxed

Timing of heart contraction: **ventricles** rebound to relaxed shape (diastole), and **semi-lunar valves** close. Both **atria**(singular: atrium) fill with blood coming from pulmonary and systemic circulations.Pressure rises in the atria and blood begins to move into the **ventricles**.The atria then contract, forcing more blood into the ventricles. There is a pause, then ventricles contract. This raises ventricle pressure, **atrio-ventricular(AV) valves** shut and semi-lunar valves open, forcing blood from the left ventricle into the major arteries and from the right ventricle into the **aorta**.

Control for this action doesn't rely on nervous stimulation, has intrinsic rhythmicity, called **myogenic**. This is the case in mammal as well as in mollusk hearts. Other animals have **neurogenic** hearts that rely on nervous stimulation for heart action, originating in the **cardiac ganglion**.

The rhythmicity of mammalian heart relies on the **sino-atrial (SA)node**, or pacemaker. This is a **phylogenic** (based on evolutionary history) remnant of an early vertebrate heart that had one more chamber than modern hearts.

How the heart contracts: waves of **depolarization** start in **SA node** and spread through **atria**. Connectile tissue pauses the spread of depolarization at the atrial ventricular node. Signal continued by **bundle branches** to **lower ventricle**, begins to stimulate heart to contract. Contraction starts at bottom of heart at **heart apex**,then signals spread through heart.

Medulla (in the brain) controls autonomic nervous system. (The medulla is part of the brain, is continuous with the spinal cord, and controls involuntary actions of the body). **Sympathetic cardiac accelerator**connects to **spinal cord**, uses **norepinephrine** to signal. **Parasympathetic cardio-inhibitory center** reaches heart through **Vagus**

nerve, uses**acetylcholine** to signal. Hyperpolarizes membrane to inhibit heart contraction. (Autonomic nervous system: two parts working in contra to control from both sides.) Dominant effect here is **inhibitory**. If we cut Vagus nerve, heart rate promptly rises about 25 bpm.

31 Respiratory system

In humans and other animals, for example, the anatomical features of the respiratory system include **airways**, **lungs**, and the respiratory muscles.

Other animals, such as insects, have respiratory systems with very simple anatomical features, and in amphibians even the skin plays a vital role in gas exchange.

Plants also have respiratory systems but the directionality of gas exchange can be opposite to that in animals. The respiratory system in plants also includes anatomical features such as holes on the undersides of leaves known as **stomata**.

In mammals, the **diaphragm** divides the body cavity into the

```
abdominal cavity: contains the viscera (e.g., stomach and intestines)
```

```
thoracic cavity: contains the heart and lungs.
```

Respiratory tree: terminates in alveolus, alveoli. Respiratory bronchioles branch into alveolar ducts and into alveoli. Alveolus: microscopic air sacs, 300 million of these in human lungs. Total surface area large. Gas diffuses micrometer, very tiny distance.

Nervous System

Composed of tissues designed to integrate sensory information and direct a coordinated response to the environment.

Basic unit of the nervous system is the neuron[1], a highly specialized cell that uses both electrical and chemical processes to communicate. Neurons "listen" to sensory organs or other neurons, and can simultaneously "hear" from 1 to hundreds of inputs simultaneously. Likewise, a neuron can "talk" to other neurons or cells that can create an action, like muscle cells or glands.

Neurons are outnumbered in the nervous system by *glia*. Glia were once thought to only play a supportive role in helping neurons survive; today we know that they also are important participants in the communication process. Glial cells include astrocytes, ependymal cells, and a cell that has a macrophage like function. There are also oligodendocytes and Schwann cells that provide a myelin sheath.

1 http://en.wikipedia.org/wiki/Neuron

31.1 Neuron structure

Figure 9 Neuron

- Dendrite will receive information from other axons
- Stoma is the neuron body and contains typical cell parts including the mitochondria and nucleus. This is where neurotransmitters are synthesized.
- Axon hillock is where the cell body and the axon meet.
- Axon is surrounded by myelin.
  - Contains nodes of Ranvier

Types of axons include unipolar, bipolar, pseudopolar.

Action potentials results from excitatory stimuli received from the dendrites that results in a signal that travels down the axon. On the cellular level, there is a Na+ influx via channels causing a depolarization of the cell. Once those channels close, the slower opening K+ channels will open resulting in hyperpolarization by the cell.

Synapses

Neurotransmitters allow a chemical signal to be sent from one neuron to the other. Neurotransmitters must bypass a physical gap called the synapse. Examples of neurotransmitters include acetylcholine, epinephrine, and glutamate.

Example : Neuromuscular synapses

Circuits / Nuclei / Ganglia

31.2 Central nervous system

Central nervous system includes the brain and the spinal cord.

Brain There are four lobes of the brain.

- Frontal lobe - decision making
- Parietal lobe
- Temporal lobe
- Occipital lobe - vision

The areas of the brain are also dedicated to different functions.

- Precentral gyrus
- Postcentral gyrus

Protection The CNS is protected by three layers - the pia mater, the arachnoid mater, and the dura mater. Protection also comes from the circulation of cerebral spinal fluid (CSF). CSF helps to float the brain and also provide nutrients to both the brain and the spinal cord. CSF is synthesized from the choroid plexus of the lateral ventricles. In total, there are four ventricles - 2 lateral ventricles, a 3rd, and 4th ventricle.

31.3 Peripheral nervous system

The peripheral nervous system is broken down into two sub-systems, the somatic nervous system and the autonomic nervous system.

ANS - Autonomic Nervous System The ANS has two components - parasympathetic and sympathetic.

The sympathetic nervous system is the "fight or flight" or fright response and results in an increased heart rate, increased rate of breathing, and an elevated blood glucose level. There is also decreased digestion. The second neurotransmitter is epinephrine. In this case, the first neuron is short and the second neuron is long.

The parasympathetic nervous system is the rest and digest system.

Drugs must be able to pass the blood brain barrier to have an effect on the CNS. Drugs act by affecting the neuron and how likely it is to fire an action potential.

Stroke occurs when there is a blood clot that goes to the brain and prevent blood flow.

32 Sensory systems

Categorized by

1. **nature of stimulus**, such as **mechanical**, **chemical**or **light** stimulus, and
2. **where stimulus received**, such as **outside** (**exteroceptors**,such as the **eye** and **skin temp receptors**) or **inside**body (**interoceptors**, such as **blood body temperature receptors**).

Transduction of sensory input into signal. Means to "carry across",signal transduced, or carried, from environment into nervous signal.

Three sensory processes we cover

1. **taste** and **smell** (chemoreception)
2. **gravity** and **movement**
3. **light**

32.1 Taste and smell (chemoreception)

Found in mammal nose and mouth, fly feet, fish bodies, moth antennae.

Papilla: bumps on tongue, contain taste buds down between. Sweet, sour, salty and bitter. Some act directly by ion channels, others act indirectly. Other "taste" sensations really smell.

Smell: received in **nasopharynx**. Airborne molecules go into solution on moist epithelial surface of nasal passage. Approximately 1000 genescode for sensory neuron receptors. "Fried onions" odor not one receptor but a mixture of many odors registered in our mind as one. Very sensitive, habituates rapidly (don't notice a smell after a bit). Odor sensation has relatively unfiltered root to higher brain centers.

Snakes more chemosensory focused than us.

32.2 Response to gravity and movement

Registered in **inner ear**. Three semicircular canals loop in three planes at right angles to each other, responsible for transduction of movement messages. Method: hair cells

deformed by gelatinous membrane. **Vestibular apparatus**, gives us perception of gravity and movement. Due to **physical response**, not chemical binding.

Cochlea: bony, coil shaped part of inner ear, where hearing occurs.

Sound enters through **auditory canal**, vibrates **tympanic membrane**,moving three bones of middle ear (**malleus**, **incus**, and **stapes**)against oval window opening in front of cochlea. Cochlea has three fluid filled ducts, one of these the **organ of Corti**. Sound waves in air go to vibration in organ of Corti; fluid tickles **hair cells**, which register the movement along **basilar membrane** in cochlea. Different sound frequencies move different portions of basilar membrane. Hearing loss due to loss of hair cells.Humans normally smell more than 300 odors in a day(Facts and Truth).

Transduction of sound accomplished throgh **physical deformation**,not chemical binding.

32.3 Vision

Light enters **pupil*, focused by** *lens**onto** retina.*

Sclera: hardened part behind retina.

Optic nerves and **neurons** attached to retina. Blind spot where optic nerve attaches, has no receptors.

Two types of photoreceptors

1. **rods** - black and white low-light vision, 100 million in each retina in humans.

2. **cones** - color vision, work best under better illumination. 3 million in each retina.

Fovia: region of most acute vision, has most of the cones, few rods.

Transduction process of light to signal a molecular change, to light absorbing molecule called photopigment. Located in outer parts of rods and cones in pigment discs. The rod photopigment is called **rhodopsin**,cone has three photopigments, called **photopsins**. This molecular change initiates pathways to result in action potential in downstream neuron leading to vision center in brain.**Parul Godika**

Each of the three photopsins has a different peak of sensitivity: **blue**,**green** or **red**, and changes **isometric form** (from cisto trans) based on light from a particular wavelength range. Color blindness:inherited lack of one or more types of these cones. Gene carried on X chromosome, therefore more common in men than women.

32.4 Homeostasis

Is a very important part of everyone's and everything's lives. Defined as **dynamic constancy of internal environment**, maintenance of a relatively stable environment inside an organism usually involving **feedback regulation**.

Homeostasis is maintained in face of

1. a **varying external environment**, or
2. a **non-ideal, constant external environment** (as with the penguin).

Deals with **temperature**, **pH**, **chemical concentrations**,**pressure**, **oxygen levels**.

Occurs through **negative feedback loops**.

Various forms: simple thermostat in house turns off heater when above a certain temperature and on when below a certain temperature Involves **stimulus**, **sensor**,**integrating center**, **effector** and **response**.

More efficient control has **two sensors** and **two effectors**. Can be antagonistic to each other, such as, one cools, the other heats.

Precise control through **proportional control**, not all-or-none, furnace comes on a little bit if the house a bit cold. Examples in humans: vasoconstriction, change in metabolic rate, shivering. Physiological responses for high body temp: blood goes to body surface, sweating, behavioral changes (get out of sun).

Positive feedback loop: effector increases deviation from set point. Amplifies reaction. Like blood clotting process, uterine contraction during childbirth. Negative feedback must exist at some point for control.

32.5 Osmotic environments and regulations

1. **Marine invertebrates**
   a) fully marine invertebrates (not intertidal or estuarine) osmoconformers (set internal environment same as environment, no net flow of ions) in a stenohaline (narrow non-changing salt level) environment
   b) Coastal, intertidal, estuarine (ion levels fluctuate) invertebrates.Partly osmoconfomers, partly osmoregulators in a euryhaline (wide salt level variation) environment (ex: shore crab, regulates sometimes when salt levels in environment get real low).
2. **Freshwater animals**. Here, environment has lower solute concentrations than do living organisms so water tends to flow in and solutes out.
   a) Freshwater fish (bony) dilute urine, and gills actively take up ions (NaCl)
   b) Freshwater invertebrates: same situation as freshwater fish but with different structures
   c) Freshwater amphibians: active uptake of salts across their skin
3. **Marine fishes**: Here the environment has a higher solute concentration than does the organism so water tends to flow out and ions in.
   a) Bony fishes: actively secrete salts (NaCl) across gills, absorb water across gut wall, their kidney (unlike mammalian kidney) is unable to generate concentrated urine so glomerulus is reduced, active tubular secretion of $MgSO_4$
   b) cartilaginous fishes (and coelacanth): blood retains urea and trimethylamineoxide to increase its osmolality to that of seawater

4. **Terrestrial animals**: here problem is loss of water to a drier environment, and regulation of salt levels.
   a) water loss adaptations
   b) concentrated exception of salts and nitrogenous wastes

Hypoosmotic: having less osmotic potential than nearby fluid

Hyperosmotic: having more osmotic potential than nearby fluid

Isoosmotic: having equal osmotic potential than nearby fluid

Glomerulus: reduces volume of kidney

Fish started in salt water, spread to fresh water, later reinvaded salt-water environment.

Terrestrial animal water sources:

1. drinking

2. moist foods

3. from breakdown of metabolic molecules like fats. (Desert kangaroo rats get 90% of their water from metabolism.)

Secretion of nitrogenous wastes: from metabolism of amino acids, amino group has to be removed in one of three basically interchangeable chemical forms:

1. **ammonia** (aquatic life)

2. **urea** (mammals)

3. **uric acid** (birds)

Ammonia very toxic, soluble, and cheap to produce. Easy to expel for bony fishes.

Urea: low toxicity, good solubility, more costly to lose as it contains other groups on it. Must be released in solution, water cost.

Uric acid (white part of bird poo) low toxicity, insoluble, secreted with little water loss, more costly side groups lost than the others.

Mammalian kidney: Structure: fist-sized organ in lower back. About 1/5 of blood from aorta at any time is passing through kidneys. Blood passes through kidney many times a day.

Nephron: structural and functional unit of kidney.

Bowmans capsule: funnel-like opening, contains primary filter, the **glomerulus**.

Proximal convoluted tubule: receives stuff from Bowmans capsule.

Loop of Henle: descends and ascends.

Vasa recta: capillaries that surround the Loop of Henle.

Glomerulus: main filter of the **nephron**, located within the **Bowman's capsule**

Kidney properties and processes important to its function

1. **Active transport of solutes** from one fluid to another against a concentration gradient, Na^+ actively transported out of filtrate by cells of the thick ascending loop of Henley into the interstitial fluid

2. **Passive movement of solutes** and water from one fluid to another(down a concentration gradient), movement of water and NaCl out of descending loop of Henley into interstitial fluid.

3. **Differential permeability of cells** in different regions of the nephron to movement of water and solutes, ascending thick look is impermeable to water, descending portion is permeable to water

4. **Hormonal control of that permeability**, antidiuretic hormone(ADH) increases permeability of collecting due to water, resulting in reduced volume of filtrate and thus more concentrated urine.

5. **Increasing solute concentration in the interstitial fluid of the kidney**, from the cortex to the deepest medulla, maintained by a countercurrent multiplier mechanism

33 Additional material

- Francis Crick[1] *chemist and molecular biologist, discovered structure of DNA molecule*
- Charles Darwin[2] *the father of the science of evolutionary biology*
- Richard Dawkins[3] *zoologist and biology populariser*
- Stephen Jay Gould[4] *paleontologist and science populariser*
- J.B.S. Haldane[5] *geneticist and evolutionary biologist, founded population genetics and the modern synthesis*
- Bill Hamilton[6] *formulated theory of inclusive fitness and kin selection*
- Thomas Huxley[7] *"Darwin's Bulldog", early evolutionary biologist and science populariser*
- Lynn Margulis[8] *introduced the theory of eukaryotic cell origin through endosymbiosis*
- Barbara McClintock[9] *geneticist and molecular biologist, discovered transposons*
- Gregor Mendel[10] *discovered the basic rules of heredity*
- Ernst Mayr[11] *evolutionary biologist and science populariser*
- Mark Ridley[12] *science populariser*
- Fred Sanger[13] *founder of DNA and protein sequencing techniques*
- John Maynard Smith[14] *evolutionary biologist and science populariser*
- Alfred Russel Wallace[15] *evolutionary biologist*
- James Watson[16] *molecular biologist, discovered structure of DNA molecule*
- Edward Wilson[17] *founded "sociobiology"*

This book is intended as a compilation of biographies describing the lives and work of influential biologists.

1 http://en.wikipedia.org/wiki/Francis%20Crick
2 http://en.wikibooks.org/wiki/General%20Biology%2FGallery%20of%20Biologists%2FCharles%20Darwin
3 http://en.wikipedia.org/wiki/Richard%20Dawkins
4 http://en.wikipedia.org/wiki/Stephen%20Jay%20Gould
5 http://en.wikipedia.org/wiki/J.B.S.%20Haldane
6 http://en.wikipedia.org/wiki/William%20Hamilton
7 http://en.wikipedia.org/wiki/Thomas%20Huxley
8 http://en.wikipedia.org/wiki/Lynn%20Margulis
9 http://en.wikipedia.org/wiki/Barbara%20McClintock
10 http://en.wikipedia.org/wiki/Gregor%20Mendel
11 http://en.wikipedia.org/wiki/Ernst%20Mayr
12 http://en.wikipedia.org/wiki/Mark%20Ridley
13 http://en.wikipedia.org/wiki/Fred%20Sanger
14 http://en.wikipedia.org/wiki/John%20Maynard%20Smith
15 http://en.wikipedia.org/wiki/Alfred%20Russel%20Wallace
16 http://en.wikipedia.org/wiki/James%20Watson
17 http://en.wikipedia.org/wiki/Edward%20O.%20Wilson

33.1 External Links

- EvoWiki: List of Biologists[18]

18 http://wiki.cotch.net/index.php/List_of_biologists

34 Glossary

- **Autotroph**[1]: an organism which can make its own energy
- **Cell**[2]: Fundamental structural unit of all living things
- **Ether**[3]:
- **Eukaryote**[4]: an organism[5] with a nucleus
- **Exoenzyme**[6]: an enzyme used to break down organic molecules[7] outside the body
- **Glycerol**[8]:
- **Heterotroph**[9]: an organism which can not make its own energy
- **Hydrocarbon**[10]: an organic compound that contains carbon[11] and hydrogen[12] only.
- **Lipid**[13]: fatty acid[14] esters[15] which form the basis of cell membranes
- **Nucleus**[16]: Membrane[17]-bound organelle[18] which contains the chromosomes[19]
- **Prokaryote**[20]: an organism with no nucleus
- **Seed**:
- **Flower**:
- **Tracheid**:
- **Haploid**: A cell with a single set of chromosomes (23 in humans), in humans this is usually in gametes. This is commonly represented by n.
- **Diploid**: A cell with two sets of chromosomes (46 in humans). This is commonly represented by 2n.
- **Sporangium**:

1 http://en.wikibooks.org/wiki/Autotroph
2 http://en.wikibooks.org/wiki/Cell
3 http://en.wikibooks.org/wiki/Ether
4 http://en.wikibooks.org/wiki/Eukaryote
5 http://en.wikibooks.org/wiki/organism
6 http://en.wikibooks.org/wiki/Exoenzyme
7 http://en.wikibooks.org/wiki/organic%20molecule
8 http://en.wikibooks.org/wiki/Glycerol
9 http://en.wikibooks.org/wiki/Heterotroph
10 http://en.wikibooks.org/wiki/Hydrocarbon
11 http://en.wikibooks.org/wiki/Carbon
12 http://en.wikibooks.org/wiki/Hydrogen
13 http://en.wikibooks.org/wiki/Lipid
14 http://en.wikibooks.org/wiki/fatty%20acid
15 http://en.wikibooks.org/wiki/ester
16 http://en.wikibooks.org/wiki/Cell%20nucleus
17 http://en.wikibooks.org/wiki/Membrane
18 http://en.wikibooks.org/wiki/organelle
19 http://en.wikibooks.org/wiki/chromosome
20 http://en.wikibooks.org/wiki/Prokaryote

The majority of the modules making up this book are based on notes very generously donated by Paul Doerder, Ph.D.[21] and Ralph Gibson, Ph.D.[22] both currently of the Cleveland State University[23].

The book was initiated by Karl Wick[24], who donated many of his own class notes for other modules, and who is fleshing out the outline format of Dr. Doerder's notes into text.

34.1 Users

Alsocal[25]

Darren Hess[26] MD/PhD recent grad, enjoys teaching, hopes to help work up the Nervous System Tissue section.

21 http://bgesweb.artscipub.csuohio.edu/faculty/doerder.htm
22 http://bgesweb.artscipub.csuohio.edu/faculty/gibson.htm
23 http://www.csuohio.edu/
24 http://en.wikibooks.org/wiki/User%3AKarl%20Wick
25 http://en.wikibooks.org/wiki/User%3AAlsocal
26 http://en.wikibooks.org/wiki/User%3Admhessmdphd

35 Contributors

Edits	User
15	Adrignola[1]
1	Alex.Szatmary[2]
72	Alsocal[3]
2	Arcanian Kastania[4]
12	Avicennasis[5]
1	Az1568[6]
1	Benlee[7]
1	Bhatiajigar406[8]
3	Bluelightstar[9]
1	Blurpeace[10]
2	Brother green[11]
1	Cnelson[12]
1	CommonsDelinker[13]
1	Dallas1278[14]
1	DavidCary[15]
10	Derbeth[16]
2	Dets65[17]
1	Dhs001[18]
8	Dirk Hünniger[19]
3	Dmhessmdphd[20]
1	Dysprosia[21]

1 http://en.wikibooks.org/w/index.php?title=User:Adrignola
2 http://en.wikibooks.org/w/index.php?title=User:Alex.Szatmary
3 http://en.wikibooks.org/w/index.php?title=User:Alsocal
4 http://en.wikibooks.org/w/index.php?title=User:Arcanian_Kastania
5 http://en.wikibooks.org/w/index.php?title=User:Avicennasis
6 http://en.wikibooks.org/w/index.php?title=User:Az1568
7 http://en.wikibooks.org/w/index.php?title=User:Benlee
8 http://en.wikibooks.org/w/index.php?title=User:Bhatiajigar406
9 http://en.wikibooks.org/w/index.php?title=User:Bluelightstar
10 http://en.wikibooks.org/w/index.php?title=User:Blurpeace
11 http://en.wikibooks.org/w/index.php?title=User:Brother_green
12 http://en.wikibooks.org/w/index.php?title=User:Cnelson
13 http://en.wikibooks.org/w/index.php?title=User:CommonsDelinker
14 http://en.wikibooks.org/w/index.php?title=User:Dallas1278
15 http://en.wikibooks.org/w/index.php?title=User:DavidCary
16 http://en.wikibooks.org/w/index.php?title=User:Derbeth
17 http://en.wikibooks.org/w/index.php?title=User:Dets65
18 http://en.wikibooks.org/w/index.php?title=User:Dhs001
19 http://en.wikibooks.org/w/index.php?title=User:Dirk_H%C3%BCnniger
20 http://en.wikibooks.org/w/index.php?title=User:Dmhessmdphd
21 http://en.wikibooks.org/w/index.php?title=User:Dysprosia

1 Ec5618[22]
2 Ervinn[23]
1 Fishpi[24]
1 Gentgeen[25]
5 Geocachernemesis[26]
1 Greentea36[27]
4 Guanabot[28]
1 Hagindaz[29]
1 Hanaa[30]
5 Herbythyme[31]
1 Iamunknown[32]
6 Idraax[33]
4 InfinityLiger[34]
1 Italienmoose[35]
2 Jaberwocky6669[36]
4 Jackaja[37]
1 Javariel[38]
1 Jclee[39]
17 Jennrulez757[40]
33 JetL011[41]
19 Jguk[42]
6 Jlee52789[43]
35 Jomegat[44]
2 Juadonsalazar[45]
1 Kamran ahmad[46]

22 http://en.wikibooks.org/w/index.php?title=User:Ec5618
23 http://en.wikibooks.org/w/index.php?title=User:Ervinn
24 http://en.wikibooks.org/w/index.php?title=User:Fishpi
25 http://en.wikibooks.org/w/index.php?title=User:Gentgeen
26 http://en.wikibooks.org/w/index.php?title=User:Geocachernemesis
27 http://en.wikibooks.org/w/index.php?title=User:Greentea36
28 http://en.wikibooks.org/w/index.php?title=User:Guanabot
29 http://en.wikibooks.org/w/index.php?title=User:Hagindaz
30 http://en.wikibooks.org/w/index.php?title=User:Hanaa
31 http://en.wikibooks.org/w/index.php?title=User:Herbythyme
32 http://en.wikibooks.org/w/index.php?title=User:Iamunknown
33 http://en.wikibooks.org/w/index.php?title=User:Idraax
34 http://en.wikibooks.org/w/index.php?title=User:InfinityLiger
35 http://en.wikibooks.org/w/index.php?title=User:Italienmoose
36 http://en.wikibooks.org/w/index.php?title=User:Jaberwocky6669
37 http://en.wikibooks.org/w/index.php?title=User:Jackaja
38 http://en.wikibooks.org/w/index.php?title=User:Javariel
39 http://en.wikibooks.org/w/index.php?title=User:Jclee
40 http://en.wikibooks.org/w/index.php?title=User:Jennrulez757
41 http://en.wikibooks.org/w/index.php?title=User:JetL011
42 http://en.wikibooks.org/w/index.php?title=User:Jguk
43 http://en.wikibooks.org/w/index.php?title=User:Jlee52789
44 http://en.wikibooks.org/w/index.php?title=User:Jomegat
45 http://en.wikibooks.org/w/index.php?title=User:Juadonsalazar
46 http://en.wikibooks.org/w/index.php?title=User:Kamran_ahmad

179	Karl Wick[47]
1	Karthik[48]
1	Kksf1979[49]
1	Kotireddyanekallu[50]
12	Lazyquasar[51]
2	Lexor[52]
3	Liblamb[53]
1	Ltellez[54]
1	MarkHudson[55]
4	Marshman[56]
1	Mathonius[57]
4	Mattb112885[58]
1	Mh7kJ[59]
7	Mike.lifeguard[60]
9	Moadeeb[61]
7	Monk[62]
1	Naryathegreat[63]
2	Neoptolemus[64]
1	NipplesMeCool[65]
1	Nmontague[66]
2	Obscureownership[67]
4	Panic2k4[68]
1	Perl[69]
1	Petemella[70]
16	QuiteUnusual[71]

47 http://en.wikibooks.org/w/index.php?title=User:Karl_Wick
48 http://en.wikibooks.org/w/index.php?title=User:Karthik
49 http://en.wikibooks.org/w/index.php?title=User:Kksf1979
50 http://en.wikibooks.org/w/index.php?title=User:Kotireddyanekallu
51 http://en.wikibooks.org/w/index.php?title=User:Lazyquasar
52 http://en.wikibooks.org/w/index.php?title=User:Lexor
53 http://en.wikibooks.org/w/index.php?title=User:Liblamb
54 http://en.wikibooks.org/w/index.php?title=User:Ltellez
55 http://en.wikibooks.org/w/index.php?title=User:MarkHudson
56 http://en.wikibooks.org/w/index.php?title=User:Marshman
57 http://en.wikibooks.org/w/index.php?title=User:Mathonius
58 http://en.wikibooks.org/w/index.php?title=User:Mattb112885
59 http://en.wikibooks.org/w/index.php?title=User:Mh7kJ
60 http://en.wikibooks.org/w/index.php?title=User:Mike.lifeguard
61 http://en.wikibooks.org/w/index.php?title=User:Moadeeb
62 http://en.wikibooks.org/w/index.php?title=User:Monk
63 http://en.wikibooks.org/w/index.php?title=User:Naryathegreat
64 http://en.wikibooks.org/w/index.php?title=User:Neoptolemus
65 http://en.wikibooks.org/w/index.php?title=User:NipplesMeCool
66 http://en.wikibooks.org/w/index.php?title=User:Nmontague
67 http://en.wikibooks.org/w/index.php?title=User:Obscureownership
68 http://en.wikibooks.org/w/index.php?title=User:Panic2k4
69 http://en.wikibooks.org/w/index.php?title=User:Perl
70 http://en.wikibooks.org/w/index.php?title=User:Petemella
71 http://en.wikibooks.org/w/index.php?title=User:QuiteUnusual

1 Ravichandar84[72]
16 Recent Runes[73]
2 Robert Horning[74]
1 Rymwoo[75]
1 SB Johnny[76]
2 SMH[77]
32 Sam Stultus[78]
3 Savh[79]
1 Shaggyjacobs[80]
2 Silkiesttie[81]
2 Smellyone[82]
1 Some P. Erson[83]
1 Starwindfury[84]
3 Steinsky[85]
2 Swift[86]
1 T.D. Migneault[87]
2 TUF-KAT[88]
1 Tannin[89]
12 Taoster[90]
1 Teaandcrumpets[91]
13 Thenub314[92]
1 Theornamentalist[93]
15 Toriber[94]
5 Uncle G[95]
1 Webaware[96]

72 http://en.wikibooks.org/w/index.php?title=User:Ravichandar84
73 http://en.wikibooks.org/w/index.php?title=User:Recent_Runes
74 http://en.wikibooks.org/w/index.php?title=User:Robert_Horning
75 http://en.wikibooks.org/w/index.php?title=User:Rymwoo
76 http://en.wikibooks.org/w/index.php?title=User:SB_Johnny
77 http://en.wikibooks.org/w/index.php?title=User:SMH
78 http://en.wikibooks.org/w/index.php?title=User:Sam_Stultus
79 http://en.wikibooks.org/w/index.php?title=User:Savh
80 http://en.wikibooks.org/w/index.php?title=User:Shaggyjacobs
81 http://en.wikibooks.org/w/index.php?title=User:Silkiesttie
82 http://en.wikibooks.org/w/index.php?title=User:Smellyone
83 http://en.wikibooks.org/w/index.php?title=User:Some_P._Erson
84 http://en.wikibooks.org/w/index.php?title=User:Starwindfury
85 http://en.wikibooks.org/w/index.php?title=User:Steinsky
86 http://en.wikibooks.org/w/index.php?title=User:Swift
87 http://en.wikibooks.org/w/index.php?title=User:T.D._Migneault
88 http://en.wikibooks.org/w/index.php?title=User:TUF-KAT
89 http://en.wikibooks.org/w/index.php?title=User:Tannin
90 http://en.wikibooks.org/w/index.php?title=User:Taoster
91 http://en.wikibooks.org/w/index.php?title=User:Teaandcrumpets
92 http://en.wikibooks.org/w/index.php?title=User:Thenub314
93 http://en.wikibooks.org/w/index.php?title=User:Theornamentalist
94 http://en.wikibooks.org/w/index.php?title=User:Toriber
95 http://en.wikibooks.org/w/index.php?title=User:Uncle_G
96 http://en.wikibooks.org/w/index.php?title=User:Webaware

3 Whiteknight[97]
8 Xania[98]
1 Yada[99]
1 Zarius[100]

97 http://en.wikibooks.org/w/index.php?title=User:Whiteknight
98 http://en.wikibooks.org/w/index.php?title=User:Xania
99 http://en.wikibooks.org/w/index.php?title=User:Yada
100 http://en.wikibooks.org/w/index.php?title=User:Zarius

List of Figures

- GFDL: Gnu Free Documentation License. http://www.gnu.org/licenses/fdl.html
- cc-by-sa-3.0: Creative Commons Attribution ShareAlike 3.0 License. http://creativecommons.org/licenses/by-sa/3.0/
- cc-by-sa-2.5: Creative Commons Attribution ShareAlike 2.5 License. http://creativecommons.org/licenses/by-sa/2.5/
- cc-by-sa-2.0: Creative Commons Attribution ShareAlike 2.0 License. http://creativecommons.org/licenses/by-sa/2.0/
- cc-by-sa-1.0: Creative Commons Attribution ShareAlike 1.0 License. http://creativecommons.org/licenses/by-sa/1.0/
- cc-by-2.0: Creative Commons Attribution 2.0 License. http://creativecommons.org/licenses/by/2.0/
- cc-by-2.0: Creative Commons Attribution 2.0 License. http://creativecommons.org/licenses/by/2.0/deed.en
- cc-by-2.5: Creative Commons Attribution 2.5 License. http://creativecommons.org/licenses/by/2.5/deed.en
- cc-by-3.0: Creative Commons Attribution 3.0 License. http://creativecommons.org/licenses/by/3.0/deed.en
- GPL: GNU General Public License. http://www.gnu.org/licenses/gpl-2.0.txt
- LGPL: GNU Lesser General Public License. http://www.gnu.org/licenses/lgpl.html
- PD: This image is in the public domain.
- ATTR: The copyright holder of this file allows anyone to use it for any purpose, provided that the copyright holder is properly attributed. Redistribution, derivative work, commercial use, and all other use is permitted.
- EURO: This is the common (reverse) face of a euro coin. The copyright on the design of the common face of the euro coins belongs to the European Commission. Authorised is reproduction in a format without relief (drawings, paintings, films) provided they are not detrimental to the image of the euro.
- LFK: Lizenz Freie Kunst. http://artlibre.org/licence/lal/de
- CFR: Copyright free use.

- EPL: Eclipse Public License. `http://www.eclipse.org/org/documents/epl-v10.php`

Copies of the GPL, the LGPL as well as a GFDL are included in chapter Licenses[101]. Please note that images in the public domain do not require attribution. You may click on the image numbers in the following table to open the webpage of the images in your webbrower.

101 Chapter 36 on page 179

1		GFDL
2	User:[102]	PD
3	Unknown, *The Hornet* is no longer in publication and it is very likely for a 20-year-old artist in 1871 to have died before 1939	PD
4		PD
5		GFDL
6		GFDL
7		PD
8	Uploader, CDC	GFDL
9	LadyofHats[103]	PD

102 http://en.wikibooks.org/wiki/User%3A
103 http://en.wikibooks.org/wiki/User%3ALadyofHats

36 Licenses

36.1 GNU GENERAL PUBLIC LICENSE

Version 3, 29 June 2007

Copyright © 2007 Free Software Foundation, Inc. <http://fsf.org/>

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed. Preamble

The GNU General Public License is a free, copyleft license for software and other kinds of works.

The licenses for most software and other practical works are designed to take away your freedom to share and change the works. By contrast, the GNU General Public License is intended to guarantee your freedom to share and change all versions of a program–to make sure it remains free software for all its users. We, the Free Software Foundation, use the GNU General Public License for most of our software; it applies also to any other work released this way by its authors. You can apply it to your programs, too.

When we speak of free software, we are referring to freedom, not price. Our General Public Licenses are designed to make sure that you have the freedom to distribute copies of free software (and charge for them if you wish), that you receive source code or can get it if you want it, that you can change the software or use pieces of it in new free programs, and that you know you can do these things.

To protect your rights, we need to prevent others from denying you these rights or asking you to surrender the rights. Therefore, you have certain responsibilities if you distribute copies of the software, or if you modify it: responsibilities to respect the freedom of others.

For example, if you distribute copies of such a program, whether gratis or for a fee, you must pass on to the recipients the same freedoms that you received. You must make sure that they, too, receive or can get the source code. And you must show them these terms so they know their rights.

Developers that use the GNU GPL protect your rights with two steps: (1) assert copyright on the software, and (2) offer you this License giving you legal permission to copy, distribute and/or modify it.

For the developers' and authors' protection, the GPL clearly explains that there is no warranty for this free software. For both users' and authors' sake, the GPL requires that modified versions be marked as changed, so that their problems will not be attributed erroneously to authors of previous versions.

Some devices are designed to deny users access to install or run modified versions of the software inside them, although the manufacturer can do so. This is fundamentally incompatible with the aim of protecting users' freedom to change the software. The systematic pattern of such abuse occurs in the area of products for individuals to use, which is precisely where it is most unacceptable. Therefore, we have designed this version of the GPL to prohibit the practice for those products. If such problems arise substantially in other domains, we stand ready to extend this provision to those domains in future versions of the GPL, as needed to protect the freedom of users.

Finally, every program is threatened constantly by software patents. States should not allow patents to restrict development and use of software on general-purpose computers, but in those that do, we wish to avoid the special danger that patents applied to a free program could make it effectively proprietary. To prevent this, the GPL assures that patents cannot be used to render the program non-free.

The precise terms and conditions for copying, distribution and modification follow. TERMS AND CONDITIONS 0. Definitions.

"This License" refers to version 3 of the GNU General Public License.

"Copyright" also means copyright-like laws that apply to other kinds of works, such as semiconductor masks.

"The Program" refers to any copyrightable work licensed under this License. Each licensee is addressed as "you". "Licensees" and "recipients" may be individuals or organizations.

To "modify" a work means to copy from or adapt all or part of the work in a fashion requiring copyright permission, other than the making of an exact copy. The resulting work is called a "modified version" of the earlier work or a work "based on" the earlier work.

A "covered work" means either the unmodified Program or a work based on the Program.

To "propagate" a work means to do anything with it that, without permission, would make you directly or secondarily liable for infringement under applicable copyright law, except executing it on a computer or modifying a private copy. Propagation includes copying, distribution (with or without modification), making available to the public, and in some countries other activities as well.

To "convey" a work means any kind of propagation that enables other parties to make or receive copies. Mere interaction with a user through a computer network, with no transfer of a copy, is not conveying.

An interactive user interface displays "Appropriate Legal Notices" to the extent that it includes a convenient and prominently visible feature that (1) displays an appropriate copyright notice, and (2) tells the user that there is no warranty for the work (except to the extent that warranties are provided), that licensees may convey the work under this License, and how to view a copy of this License. If the interface presents a list of user commands or options, such as a menu, a prominent item in the list meets this criterion. 1. Source Code.

The "source code" for a work means the preferred form of the work for making modifications to it. "Object code" means any non-source form of a work.

A "Standard Interface" means an interface that either is an official standard defined by a recognized standards body, or, in the case of interfaces specified for a particular programming language, one that is widely used among developers working in that language.

The "System Libraries" of an executable work include anything, other than the work as a whole, that (a) is included in the normal form of packaging a Major Component, but which is not part of that Major Component, and (b) serves only to enable use of the work with that Major Component, or to implement a Standard Interface for which an implementation is available to the public in source code form. A "Major Component", in this context, means a major essential component (kernel, window system, and so on) of the specific operating system (if any) on which the executable work runs, or a compiler used to produce the work, or an object code interpreter used to run it.

The "Corresponding Source" for a work in object code form means all the source code needed to generate, install, and (for an executable work) run the object code and to modify the work, including scripts to control those activities. However, it does not include the work's System Libraries, or general-purpose tools or generally available free programs which are used unmodified in performing those activities but which are not part of the work. For example, Corresponding Source includes interface definition files associated with source files for the work, and the source code for shared libraries and dynamically linked subprograms that the work is specifically designed to require, such as by intimate data communication or control flow between those subprograms and other parts of the work.

The Corresponding Source need not include anything that users can regenerate automatically from other parts of the Corresponding Source.

The Corresponding Source for a work in source code form is that same work. 2. Basic Permissions.

All rights granted under this License are granted for the term of copyright on the Program, and are irrevocable provided the stated conditions are met. This License explicitly affirms your unlimited permission to run the unmodified Program. The output from running a covered work is covered by this License only if the output, given its content, constitutes a covered work. This License acknowledges your rights of fair use or other equivalent, as provided by copyright law.

You may make, run and propagate covered works that you do not convey, without conditions so long as your license otherwise remains in force. You may convey covered works to others for the sole purpose of having them make modifications exclusively for you, or provide you with facilities for running those works, provided that you comply with the terms of this License in conveying all material for which you do not control copyright. Those thus making or running the covered works for you must do so exclusively on your behalf, under your direction and control, on terms that prohibit them from making any copies of your copyrighted material outside their relationship with you.

Conveying under any other circumstances is permitted solely under the conditions stated below. Sublicensing is not allowed; section 10 makes it unnecessary. 3. Protecting Users' Legal Rights From Anti-Circumvention Law.

No covered work shall be deemed part of an effective technological measure under any applicable law fulfilling obligations under article 11 of the WIPO copyright treaty adopted on 20 December 1996, or similar laws prohibiting or restricting circumvention of such measures.

When you convey a covered work, you waive any legal power to forbid circumvention of technological measures to the extent such circumvention is effected by exercising rights under this License with respect to the covered work, and you disclaim any intention to limit operation or modification of the work as a means of enforcing, against the work's users, your or third parties' legal rights to forbid circumvention of technological measures. 4. Conveying Verbatim Copies.

You may convey verbatim copies of the Program's source code as you receive it, in any medium, provided that you conspicuously and appropriately publish on each copy an appropriate copyright notice; keep intact all notices stating that this License and any non-permissive terms added in accord with section 7 apply to the code; keep intact all notices of the absence of any warranty; and give all recipients a copy of this License along with the Program. You may charge any price or no price for each copy that you convey, and you may offer support or warranty protection for a fee. 5. Conveying Modified Source Versions.

You may convey a work based on the Program, or the modifications to produce it from the Program, in the form of source code under the terms of section 4, provided that you also meet all of these conditions:

* a) The work must carry prominent notices stating that you modified it, and giving a relevant date. * b) The work must carry prominent notices stating that it is released under this License and any conditions added under section 7. This requirement modifies the requirement in section 4 to "keep intact all notices". * c) You must license the entire work, as a whole, under this License to anyone who comes into possession of a copy. This License will therefore apply, along with any applicable section 7 additional terms, to the whole of the work, and all its parts, regardless of how they are packaged. This License gives no permission to license the work in any other way, but it does not invalidate such permission if you have separately received it. * d) If the work has interactive user interfaces, each must display Appropriate Legal Notices; however, if the Program has interactive interfaces that do not display Appropriate Legal Notices, your work need not make them do so.

A compilation of a covered work with other separate and independent works, which are not by their nature extensions of the covered work, and which are not combined with it such as to form a larger program, in or on a volume of a storage or distribution medium, is called an "aggregate" if the compilation and its resulting copyright are not used to limit the access or legal rights of the compilation's users beyond what the individual works permit. Inclusion of a covered work in an aggregate does not cause this License to apply to the other parts of the aggregate. 6. Conveying Non-Source Forms.

You may convey a covered work in object code form under the terms of sections 4 and 5, provided that you also convey the machine-readable Corresponding Source under the terms of this License, in one of these ways:

* a) Convey the object code in, or embodied in, a physical product (including a physical distribution medium), accompanied by the Corresponding Source fixed on a durable physical medium customarily used for software interchange. * b) Convey the object code in, or embodied in, a physical product (including a physical distribution medium), accompanied by a written offer, valid for at least three years and valid for as long as you offer spare parts or customer support for that product model, to give anyone who possesses the object code either (1) a copy of the Corresponding Source for all the software in the product that is covered by this License, on a durable physical medium customarily used for software interchange, for a price no more than your reasonable cost of physically performing this conveying of source, or (2) access to copy the Corresponding Source from a network server at no charge. * c) Convey individual copies of the object code with a copy of the written offer to provide the Corresponding Source. This alternative is allowed only occasionally and noncommercially, and only if you received the object code with such an offer, in accord with subsection 6b. * d) Convey the object code by offering access from a designated place (gratis or for a charge), and offer equivalent access to the Corresponding Source in the same way through the same place at no further charge. You need not require recipients to copy the Corresponding Source along with the object code. If the place to copy the object code is a network server, the Corresponding Source may be on a different server (operated by you or a third party) that supports equivalent copying facilities, provided you maintain clear directions next to the object code saying where to find the Corresponding Source. Regardless of what server hosts the Corresponding Source, you remain obligated to ensure that it is available for as long as needed to satisfy these requirements. * e) Convey the object code using peer-to-peer transmission, provided you inform other peers where the object code and Corresponding Source of the work are being offered to the general public at no charge under subsection 6d.

A separable portion of the object code, whose source code is excluded from the Corresponding Source as a System Library, need not be included in conveying the object code work.

A "User Product" is either (1) a "consumer product", which means any tangible personal property which is normally used for personal, family, or household purposes, or (2) anything designed or sold for incorporation into a dwelling. In determining whether a product is a consumer product, doubtful cases shall be resolved in favor of coverage. For a particular product received by a particular user, "normally used" refers to a typical or common use of that class of product, regardless of the status of the particular user or of the way in which the particular user actually uses, or expects or is expected to use, the product. A product is a consumer product regardless of whether the product has substantial commercial, industrial or non-consumer uses, unless such uses represent the only significant mode of use of the product.

"Installation Information" for a User Product means any methods, procedures, authorization keys, or other information required to install and execute modified versions of a covered work in that User Product from a modified version of its Corresponding Source. The information must suffice to ensure that the continued functioning of the modified object code is in no case prevented or interfered with solely because modification has been made.

If you convey an object code work under this section in, or with, or specifically for use in, a User Product, and the conveying occurs as part of a transaction in which the right of possession and use of the User Product is transferred to the recipient in perpetuity or for a fixed term (regardless of how the transaction is characterized), the Corresponding Source conveyed under this section must be accompanied by the Installation Information. But this requirement does not apply if neither you nor any third party retains the ability to install modified object code on the User Product (for example, the work has been installed in ROM).

The requirement to provide Installation Information does not include a requirement to continue to provide support service, warranty, or updates for a work that has been modified or installed by the recipient, or for the User Product in which it has been modified or installed. Access to a network may be denied when the modification itself materially and adversely affects the operation of the network or violates the rules and protocols for communication across the network.

Corresponding Source conveyed, and Installation Information provided, in accord with this section must be in a format that is publicly documented (and with an implementation available to the public in source code form), and must require no special password or key for unpacking, reading or copying. 7. Additional Terms.

"Additional permissions" are terms that supplement the terms of this License by making exceptions from one or more of its conditions. Additional permissions that are applicable to the entire Program shall be treated as though they were included in this License, to the extent that they are valid under applicable law. If additional permissions apply only to part of the Program, that part may be used separately under those permissions, but the entire Program remains governed by this License without regard to the additional permissions.

When you convey a copy of a covered work, you may at your option remove any additional permissions from that copy, or from any part of it. (Additional permissions may be written to require their own removal in certain cases when you modify the work.) You may place additional permissions on material, added by you to a covered work, for which you have or can give appropriate copyright permission.

Notwithstanding any other provision of this License, for material you add to a covered work, you may (if authorized by the copyright holders of that material) supplement the terms of this License with terms:

* a) Disclaiming warranty or limiting liability differently from the terms of sections 15 and 16 of this License; or * b) Requiring preservation of specified reasonable legal notices or author attributions in that material or in the Appropriate Legal Notices displayed by works containing it; or * c) Prohibiting misrepresentation of the origin of that material, or requiring that modified versions of such material be marked in reasonable ways as different from the original version; or * d) Limiting the use for publicity purposes of names of licensors or authors of the material; or * e) Declining to grant rights under trademark law for use of some trade names, trademarks, or service marks; or * f) Requiring indemnification of licensors and authors of that material by anyone who conveys the material (or modified versions of it) with contractual assumptions of liability to the recipient, for any liability that these contractual assumptions directly impose on those licensors and authors.

All other non-permissive additional terms are considered "further restrictions" within the meaning of section 10. If the Program as you received it, or any part of it, contains a notice stating that it is governed by this License along with a term that is a further restriction, you may remove that term. If a license document contains a further restriction but permits relicensing or conveying under this License, you may add to a covered work material governed by the terms of that license document, provided that the further restriction does not survive such relicensing or conveying.

If you add terms to a covered work in accord with this section, you must place, in the relevant source files, a statement of the additional terms that apply to those files, or a notice indicating where to find the applicable terms.

Additional terms, permissive or non-permissive, may be stated in the form of a separately written license, or stated as exceptions; the above requirements apply either way. 8. Termination.

You may not propagate or modify a covered work except as expressly provided under this License. Any attempt otherwise to propagate or modify it is void, and will automatically terminate your rights under this License (including any patent licenses granted under the third paragraph of section 11).

However, if you cease all violation of this License, then your license from a particular copyright holder is reinstated (a) provisionally, unless and until the copyright holder explicitly and finally terminates your license, and (b) permanently, if the copyright holder fails to notify you of the violation by some reasonable means prior to 60 days after the cessation.

Moreover, your license from a particular copyright holder is reinstated permanently if the copyright holder notifies you of the violation by some reasonable means, this is the first time you have received notice of violation of this License (for any work)

from that copyright holder, and you cure the violation prior to 30 days after your receipt of the notice.

Termination of your rights under this section does not terminate the licenses of parties who have received copies or rights from you under this License. If your rights have been terminated and not permanently reinstated, you do not qualify to receive new licenses for the same material under section 10. 9. Acceptance Not Required for Having Copies.

You are not required to accept this License in order to receive or run a copy of the Program. Ancillary propagation of a covered work occurring solely as a consequence of using peer-to-peer transmission to receive a copy likewise does not require acceptance. However, nothing other than this License grants you permission to propagate or modify any covered work. These actions infringe copyright if you do not accept this License. Therefore, by modifying or propagating a covered work, you indicate your acceptance of this License to do so. 10. Automatic Licensing of Downstream Recipients.

Each time you convey a covered work, the recipient automatically receives a license from the original licensors, to run, modify and propagate that work, subject to this License. You are not responsible for enforcing compliance by third parties with this License.

An "entity transaction" is a transaction transferring control of an organization, or substantially all assets of one, or subdividing an organization, or merging organizations. If propagation of a covered work results from an entity transaction, each party to that transaction who receives a copy of the work also receives whatever licenses to the work the party's predecessor in interest had or could give under the previous paragraph, plus a right to possession of the Corresponding Source of the work from the predecessor in interest, if the predecessor has it or can get it with reasonable efforts.

You may not impose any further restrictions on the exercise of the rights granted or affirmed under this License. For example, you may not impose a license fee, royalty, or other charge for exercise of rights granted under this License, and you may not initiate litigation (including a cross-claim or counterclaim in a lawsuit) alleging that any patent claim is infringed by making, using, selling, offering for sale, or importing the Program or any portion of it. 11. Patents.

A "contributor" is a copyright holder who authorizes use under this License of the Program or a work on which the Program is based. The work thus licensed is called the contributor's "contributor version".

A contributor's "essential patent claims" are all patent claims owned or controlled by the contributor, whether already acquired or hereafter acquired, that would be infringed by some manner, permitted by this License, of making, using, or selling its contributor version, but do not include claims that would be infringed only as a consequence of further modification of the contributor version. For purposes of this definition, "control" includes the right to grant patent sublicenses in a manner consistent with the requirements of this License.

Each contributor grants you a non-exclusive, worldwide, royalty-free patent license under the contributor's essential patent claims, to make, use, sell, offer for sale, import and otherwise run, modify and propagate the contents of its contributor version.

In the following three paragraphs, a "patent license" is any express agreement or commitment, however denominated, not to enforce a patent (such as an express permission to practice a patent or covenant not to sue for patent infringement). To "grant" such a patent license to a party means to make such an agreement or commitment not to enforce a patent against the party.

If you convey a covered work, knowingly relying on a patent license, and the Corresponding Source of the work is not available for anyone to copy, free of charge and under the terms of this License, through a publicly available network server or other readily accessible means, then you must either (1) cause the Corresponding Source to be so available, or (2) arrange to deprive yourself of the benefit of the patent license for this particular work, or (3) arrange, in a manner consistent with the requirements of this License, to extend the patent license to downstream recipients. "Knowingly relying" means you have actual knowledge that, but for the patent license, your conveying the covered work in a country, or your recipient's use of the covered work in a country, would infringe one or more identifiable patents in that country that you have reason to believe are valid.

If, pursuant to or in connection with a single transaction or arrangement, you convey, or propagate by procuring conveyance of, a covered work, and grant a patent license to some of the parties receiving the covered work authorizing them to use, propagate, modify or convey a specific copy of the covered work, then the patent license you grant is automatically extended to all recipients of the covered work and works based on it.

A patent license is "discriminatory" if it does not include within the scope of its coverage, prohibits the exercise of, or is conditioned on the non-exercise of one or more of the rights that are specifically granted under this License. You may not convey a covered work if you are a party to an arrangement with a third party that is in the business of distributing software, under which you make payment to the third party based on the extent of your activity of conveying the work, and under which the third party grants, to any of the parties who would receive the covered work from you, a discriminatory patent license (a) in connection with copies of the covered work conveyed by you (or copies made from those copies), or (b) primarily for and in connection with specific products or compilations that contain the covered work, unless you entered into that arrangement, or that patent license was granted, prior to 28 March 2007.

Nothing in this License shall be construed as excluding or limiting any implied license or other defenses to infringement that may otherwise be available to you under applicable patent law. 12. No Surrender of Others' Freedom.

If conditions are imposed on you (whether by court order, agreement or otherwise) that contradict the conditions of this License, they do not excuse you from the conditions of this License. If you cannot convey a covered work so as to satisfy simultaneously your obligations under this License and any other pertinent obligations, then as a consequence you may not convey it at all. For example, if you agree to terms that obligate you to collect a royalty for further conveying from those to whom you convey the Program, the only way you could satisfy both those terms and this License would be to refrain entirely from conveying the Program. 13. Use with the GNU Affero General Public License.

Notwithstanding any other provision of this License, you have permission to link or combine any covered work with a work licensed under version 3 of the GNU Affero General Public License into a single combined work, and to convey the resulting work. The terms of this License will continue to apply to the part which is the covered work, but the special requirements of the GNU Affero General Public License, section 13, concerning interaction through a network will apply to the combination as such. 14. Revised Versions of this License.

The Free Software Foundation may publish revised and/or new versions of the GNU General Public License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns.

Each version is given a distinguishing version number. If the Program specifies that a certain numbered version of the GNU General Public License "or any later version" applies to it, you have the option of following the terms and conditions either of that numbered version or of any later version published by the Free Software Foundation. If the Program does not specify a version number of the GNU General Public License, you may choose any version ever published by the Free Software Foundation.

If the Program specifies that a proxy can decide which future versions of the GNU General Public License can be used, that proxy's public statement of acceptance of a version permanently authorizes you to choose that version for the Program.

Later license versions may give you additional or different permissions. However, no additional obligations are imposed on any author or copyright holder as a result of your choosing to follow a later version. 15. Disclaimer of Warranty.

THERE IS NO WARRANTY FOR THE PROGRAM, TO THE EXTENT PERMITTED BY APPLICABLE LAW. EXCEPT WHEN OTHERWISE STATED IN WRITING THE COPYRIGHT HOLDERS AND/OR OTHER PARTIES PROVIDE THE PROGRAM "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. THE ENTIRE RISK AS TO THE QUALITY AND PERFORMANCE OF THE PROGRAM IS WITH YOU. SHOULD THE PROGRAM PROVE DEFECTIVE, YOU ASSUME THE COST OF ALL NECESSARY SERVICING, REPAIR OR CORRECTION. 16. Limitation of Liability.

IN NO EVENT UNLESS REQUIRED BY APPLICABLE LAW OR AGREED TO IN WRITING WILL ANY COPYRIGHT HOLDER, OR ANY OTHER PARTY WHO MODIFIES AND/OR CONVEYS THE PROGRAM AS PERMITTED ABOVE, BE LIABLE TO YOU FOR DAMAGES, INCLUDING ANY GENERAL, SPECIAL, INCIDENTAL OR CONSEQUENTIAL DAMAGES ARISING OUT OF THE USE OR INABILITY TO USE THE PROGRAM (INCLUDING BUT NOT LIMITED TO LOSS OF DATA OR DATA BEING RENDERED INACCURATE OR LOSSES SUSTAINED BY YOU OR THIRD PARTIES OR A FAILURE OF THE PROGRAM TO OPERATE WITH ANY OTHER PROGRAMS), EVEN IF SUCH HOLDER OR OTHER PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. 17. Interpretation of Sections 15 and 16.

If the disclaimer of warranty and limitation of liability provided above cannot be given local legal effect according to their terms, reviewing courts shall apply local law that most closely approximates an absolute waiver of all civil liability in connection with the Program, unless a warranty or assumption of liability accompanies a copy of the Program in return for a fee.

END OF TERMS AND CONDITIONS How to Apply These Terms to Your New Programs

If you develop a new program, and you want it to be of the greatest possible use to the public, the best way to achieve this is to make it free software which everyone can redistribute and change under these terms.

To do so, attach the following notices to the program. It is safest to attach them to the start of each source file to most effectively state the exclusion of warranty; and each file should have at least the "copyright" line and a pointer to where the full notice is found.

<one line to give the program's name and a brief idea of what it does.> Copyright (C) <year> <name of author>

This program is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License along with this program. If not, see <http://www.gnu.org/licenses/>.

Also add information on how to contact you by electronic and paper mail.

If the program does terminal interaction, make it output a short notice like this when it starts in an interactive mode:

<program> Copyright (C) <year> <name of author> This program comes with ABSOLUTELY NO WARRANTY; for details type 'show w'. This is free software, and you are welcome to redistribute it under certain conditions; type 'show c' for details.

The hypothetical commands 'show w' and 'show c' should show the appropriate parts of the General Public License. Of course, your program's commands might be different; for a GUI interface, you would use an "about box".

You should also get your employer (if you work as a programmer) or school, if any, to sign a "copyright disclaimer" for the program, if necessary. For more information on this, and how to apply and follow the GNU GPL, see <http://www.gnu.org/licenses/>.

The GNU General Public License does not permit incorporating your program into proprietary programs. If your program is a subroutine library, you may consider it more useful to permit linking proprietary applications with the library. If this is what you want to do, use the GNU Lesser General Public License instead of this License. But first, please read <http://www.gnu.org/philosophy/why-not-lgpl.html>.

36.2 GNU Free Documentation License

Version 1.3, 3 November 2008

Copyright © 2000, 2001, 2002, 2007, 2008 Free Software Foundation, Inc. <http://fsf.org/>

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed. 0. PREAMBLE

The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference. 1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "Document", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "you". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "Modified Version" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "Secondary Section" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "Invariant Sections" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The "Cover Texts" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A "Transparent" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called "Opaque".

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The "Title Page" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

The "publisher" means any person or entity that distributes copies of the Document to the public.

A section "Entitled XYZ" means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as "Acknowledgements", "Dedications", "Endorsements", or "History".) To "Preserve the Title" of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License. 2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies. 3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document. 4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

* A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission. * B. List on the Title

Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement. * C. State on the Title page the name of the publisher of the Modified Version, as the publisher. * D. Preserve all the copyright notices of the Document. * E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices. * F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below. * G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice. * H. Include an unaltered copy of this License. * I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence. * J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission. * K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein. * L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles. * M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version. * N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section. * O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties—for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version. 5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements". 6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document. 7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate. 8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title. 9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense, or distribute it is void, and will automatically terminate your rights under this License.

However, if you cease all violation of this License, then your license from a particular copyright holder is reinstated (a) provisionally, unless and until the copyright holder explicitly and finally terminates your license, and (b) permanently, if the copyright holder fails to notify you of the violation by some reasonable means prior to 60 days after the cessation.

Moreover, your license from a particular copyright holder is reinstated permanently if the copyright holder notifies you of the violation by some reasonable means, this is the first time you have received notice of violation of this License (for any work) from that copyright holder, and you cure the violation prior to 30 days after your receipt of the notice.

Termination of your rights under this section does not terminate the licenses of parties who have received copies or rights from you under this License. If your rights have been terminated and not permanently reinstated, receipt of a copy of some or all of the same material does not give you any rights to use it. 10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See http://www.gnu.org/copyleft/.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation. If the Document specifies that a proxy can decide which future versions of this License can be used, that proxy's public statement of acceptance of a version permanently authorizes you to choose that version for the Document. 11. RELICENSING

"Massive Multiauthor Collaboration Site" (or "MMC Site") means any World Wide Web server that publishes copyrightable works and also provides prominent facilities for anybody to edit those works. A public wiki that anybody can edit is an example of such a server. A "Massive Multiauthor Collaboration" (or "MMC") contained in the site means any set of copyrightable works thus published on the MMC site.

"CC-BY-SA" means the Creative Commons Attribution-Share Alike 3.0 license published by Creative Commons Corporation, a not-for-profit corporation with a principal place of business in San Francisco, California, as well as future copyleft versions of that license published by that same organization.

"Incorporate" means to publish or republish a Document, in whole or in part, as part of another Document.

An MMC is "eligible for relicensing" if it is licensed under this License, and if all works that were first published under this License somewhere other than this MMC, and subsequently incorporated in whole or in part into the MMC, (1) had no cover texts or invariant sections, and (2) were thus incorporated prior to November 1, 2008.

The operator of an MMC Site may republish an MMC contained in the site under CC-BY-SA on the same site at any time before August 1, 2009, provided the MMC is eligible for relicensing. ADDENDUM: How to use this License for your documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

Copyright (C) YEAR YOUR NAME. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License".

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the "with ... Texts." line with this:

with the Invariant Sections being LIST THEIR TITLES, with the Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.

36.3 GNU Lesser General Public License

GNU LESSER GENERAL PUBLIC LICENSE

Version 3, 29 June 2007

Copyright © 2007 Free Software Foundation, Inc. <http://fsf.org/>

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

This version of the GNU Lesser General Public License incorporates the terms and conditions of version 3 of the GNU General Public License, supplemented by the additional permissions listed below. 0. Additional Definitions.

As used herein, "this License" refers to version 3 of the GNU Lesser General Public License, and the "GNU GPL" refers to version 3 of the GNU General Public License.

"The Library" refers to a covered work governed by this License, other than an Application or a Combined Work as defined below.

An "Application" is any work that makes use of an interface provided by the Library, but which is not otherwise based on the Library. Defining a subclass of a class defined by the Library is deemed a mode of using an interface provided by the Library.

A "Combined Work" is a work produced by combining or linking an Application with the Library. The particular version of the Library with which the Combined Work was made is also called the "Linked Version".

The "Minimal Corresponding Source" for a Combined Work means the Corresponding Source for the Combined Work, excluding any source code for portions of the Combined Work that, considered in isolation, are based on the Application, and not on the Linked Version.

The "Corresponding Application Code" for a Combined Work means the object code and/or source code for the Application, including any data and utility programs needed for reproducing the Combined Work from the Application, but excluding the System Libraries of the Combined Work. 1. Exception to Section 3 of the GNU GPL.

You may convey a covered work under sections 3 and 4 of this License without being bound by section 3 of the GNU GPL. 2. Conveying Modified Versions.

If you modify a copy of the Library, and, in your modifications, a facility refers to a function or data to be supplied by an Application that uses the facility (other than as an argument passed when the facility is invoked), then you may convey a copy of the modified version:

* a) under this License, provided that you make a good faith effort to ensure that, in the event an Application does not supply the function or data, the facility still operates, and performs whatever part of its purpose remains meaningful, or * b) under the GNU GPL, with none of the additional permissions of this License applicable to that copy.

3. Object Code Incorporating Material from Library Header Files.

The object code form of an Application may incorporate material from a header file that is part of the Library. You may convey such object code under terms of your choice, provided that, if the incorporated material is not limited to numerical parameters, data structure layouts and accessors, or small macros, inline functions and templates (ten or fewer lines in length), you do both of the following:

* a) Give prominent notice with each copy of the object code that the Library is used in it and that the Library and its use are covered by this License. * b) Accompany the object code with a copy of the GNU GPL and this license document.

4. Combined Works.

You may convey a Combined Work under terms of your choice that, taken together, effectively do not restrict modification of the portions of the Library contained in the Combined Work and reverse engineering for debugging such modifications, if you also do each of the following:

* a) Give prominent notice with each copy of the Combined Work that the Library is used in it and that the Library and its use are covered by this License. * b) Accompany the Combined Work with a copy of the GNU GPL and this license document. * c) For a Combined Work that displays copyright notices during execution, include the copyright notice for the Library among these notices, as well as a reference directing the user to the copies of the GNU GPL and this license document. * d) Do one of the following: o 0) Convey the Minimal Corresponding Source under the terms of this License, and the Corresponding Application Code in a form suitable for, and under terms that permit, the user to recombine or relink the Application with a modified version of the Linked Version to produce a modified Combined Work, in the manner specified by section 6 of the GNU GPL for conveying Corresponding Source. o 1) Use a suitable shared library mechanism for linking with the Library. A suitable mechanism is one that (a) uses at run time a copy of the Library already present on the user's computer system, and (b) will operate properly with a modified version of the Library that is interface-compatible with the Linked Version. * e) Provide Installation Information, but only if you would otherwise be required to provide such information under section 6 of the GNU GPL, and only to the extent that such information is necessary to install and execute a modified version of the Combined Work produced by recombining or relinking the Application with a modified version of the Linked Version. (If you use option 4d0, the Installation Information must accompany the Minimal Corresponding Source and Corresponding Application Code. If you use option 4d1, you must provide the Installation Information in the manner specified by section 6 of the GNU GPL for conveying Corresponding Source.)

5. Combined Libraries.

You may place library facilities that are a work based on the Library side by side in a single library together with other library facilities that are not Applications and are not covered by this License, and convey such a combined library under terms of your choice, if you do both of the following:

* a) Accompany the combined library with a copy of the same work based on the Library, uncombined with any other library facilities, conveyed under the terms of this License. * b) Give prominent notice with the combined library that part of it is a work based on the Library, and explaining where to find the accompanying uncombined form of the same work.

6. Revised Versions of the GNU Lesser General Public License.

The Free Software Foundation may publish revised and/or new versions of the GNU Lesser General Public License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns.

Each version is given a distinguishing version number. If the Library as you received it specifies that a certain numbered version of the GNU Lesser General Public License "or any later version" applies to it, you have the option of following the terms and conditions either of that published version or of any later version published by the Free Software Foundation. If the Library as you received it does not specify a version number of the GNU Lesser General Public License, you may choose any version of the GNU Lesser General Public License ever published by the Free Software Foundation.

If the Library as you received it specifies that a proxy can decide whether future versions of the GNU Lesser General Public License shall apply, that proxy's public statement of acceptance of any version is permanent authorization for you to choose that version for the Library.

openstax™

Physics

HIGH SCHOOL

Physics

SENIOR CONTRIBUTING AUTHORS

Paul Peter Urone, California State University, Sacramento
Roger Hinrichs, State University of New York, College at Oswego

CONTRIBUTING AUTHORS

Fatih Gozuacik, Pennsylvania Department of Education
Denise Pattison, East Chambers ISD
Catherine Tabor, Northwest Early High School

openstax™

OpenStax
Rice University
6100 Main Street MS-375
Houston, Texas 77005

To learn more about OpenStax, visit https://openstax.org.
Individual print copies and bulk orders can be purchased through our website.

©2020 Texas Education Agency (TEA). This instructional material was initially created through a Texas Education Agency (TEA) initiative to provide high-quality open-source instructional materials to districts free of charge. Changes were made to the original material, including updates to art, structure, and other content updates. The original material is available at: https://www.texasgateway.org/book/tea-physics. Textbook content is licensed under a Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/) (CC BY 4.0).

For questions regarding this licensing, please contact support@openstax.org.

Trademarks
The OpenStax name, OpenStax logo, OpenStax book covers, OpenStax CNX name, OpenStax CNX logo, OpenStax Tutor name, Openstax Tutor logo, Connexions name, Connexions logo, Rice University name, and Rice University logo are not subject to the license and may not be reproduced without the prior and express written consent of Rice University.

HARDCOVER BOOK ISBN-13	**978-1-975076-51-1**
B&W PAPERBACK BOOK ISBN-13	**978-1-975076-50-4**
DIGITAL VERSION ISBN-13	**978-1-951693-21-3**
ORIGINAL PUBLICATION YEAR	**2020**

10 9 8 7 6 5 4 3 2 1

OpenStax

OpenStax provides free, peer-reviewed, openly licensed textbooks for introductory college and Advanced Placement® courses and low-cost, personalized courseware that helps students learn. A nonprofit ed tech initiative based at Rice University, we're committed to helping students access the tools they need to complete their courses and meet their educational goals.

Rice University

OpenStax, OpenStax CNX, and OpenStax Tutor are initiatives of Rice University. As a leading research university with a distinctive commitment to undergraduate education, Rice University aspires to path-breaking research, unsurpassed teaching, and contributions to the betterment of our world. It seeks to fulfill this mission by cultivating a diverse community of learning and discovery that produces leaders across the spectrum of human endeavor.

Philanthropic Support

OpenStax is grateful for our generous philanthropic partners, who support our vision to improve educational opportunities for all learners.

Laura and John Arnold Foundation
Arthur and Carlyse Ciocca Charitable Foundation
Ann and John Doerr
Bill & Melinda Gates Foundation
Girard Foundation
Google Inc.
The William and Flora Hewlett Foundation
Rusty and John Jaggers
The Calvin K. Kazanjian Economics Foundation
Charles Koch Foundation
Leon Lowenstein Foundation, Inc.
The Maxfield Foundation
Burt and Deedee McMurtry
Michelson 20MM Foundation
National Science Foundation
The Open Society Foundations
Jumee Yhu and David E. Park III
Brian D. Patterson USA-International Foundation
The Bill and Stephanie Sick Fund
Robin and Sandy Stuart Foundation
The Stuart Family Foundation
Tammy and Guillermo Treviño

openstax™

Study where you want, what you want, when you want.

When you access College Success in our web view, you can use our new online **highlighting and note-taking** features to create your own study guides.

Our books are free and flexible, forever.
Get started at **openstax.org/details/books/physics**

Access. The future of education.
openstax.org

RICE

Contents

Preface 1

CHAPTER 1
What is Physics? 5
Introduction 5
1.1 Physics: Definitions and Applications 5
1.2 The Scientific Methods 14
1.3 The Language of Physics: Physical Quantities and Units 18
Key Terms 40
Section Summary 41
Key Equations 41
Chapter Review 41
Test Prep 46

CHAPTER 2
Motion in One Dimension 53
Introduction 53
2.1 Relative Motion, Distance, and Displacement 54
2.2 Speed and Velocity 62
2.3 Position vs. Time Graphs 67
2.4 Velocity vs. Time Graphs 73
Key Terms 81
Section Summary 81
Key Equations 81
Chapter Review 82
Test Prep 86

CHAPTER 3
Acceleration 93
Introduction 93
3.1 Acceleration 93
3.2 Representing Acceleration with Equations and Graphs 99
Key Terms 109
Section Summary 109
Key Equations 109
Chapter Review 109
Test Prep 112

CHAPTER 4
Forces and Newton's Laws of Motion 115
Introduction 115
4.1 Force 116
4.2 Newton's First Law of Motion: Inertia 118

4.3 Newton's Second Law of Motion 122
4.4 Newton's Third Law of Motion 128
Key Terms 135
Section Summary 135
Key Equations 136
Chapter Review 136
Test Prep 139

CHAPTER 5
Motion in Two Dimensions 143
Introduction 143
5.1 Vector Addition and Subtraction: Graphical Methods 144
5.2 Vector Addition and Subtraction: Analytical Methods 153
5.3 Projectile Motion 162
5.4 Inclined Planes 171
5.5 Simple Harmonic Motion 178
Key Terms 185
Section Summary 185
Key Equations 186
Chapter Review 187
Test Prep 191

CHAPTER 6
Circular and Rotational Motion 197
Introduction 197
6.1 Angle of Rotation and Angular Velocity 198
6.2 Uniform Circular Motion 205
6.3 Rotational Motion 212
Key Terms 220
Section Summary 220
Key Equations 221
Chapter Review 221
Test Prep 223

CHAPTER 7
Newton's Law of Gravitation 229
Introduction 229
7.1 Kepler's Laws of Planetary Motion 229
7.2 Newton's Law of Universal Gravitation and Einstein's Theory of General Relativity 237
Key Terms 246
Section Summary 246
Key Equations 246
Chapter Review 246
Test Prep 249

CHAPTER 8
Momentum 253
Introduction 253

8.1 Linear Momentum, Force, and Impulse 254
8.2 Conservation of Momentum 259
8.3 Elastic and Inelastic Collisions 262
Key Terms 272
Section Summary 272
Key Equations 272
Chapter Review 273
Test Prep 275

CHAPTER 9
Work, Energy, and Simple Machines 279
Introduction 279
9.1 Work, Power, and the Work–Energy Theorem 280
9.2 Mechanical Energy and Conservation of Energy 285
9.3 Simple Machines 290
Key Terms 295
Section Summary 295
Key Equations 295
Chapter Review 296
Test Prep 299

CHAPTER 10
Special Relativity 305
Introduction 305
10.1 Postulates of Special Relativity 305
10.2 Consequences of Special Relativity 312
Key Terms 321
Section Summary 321
Key Equations 321
Chapter Review 322
Test Prep 324

CHAPTER 11
Thermal Energy, Heat, and Work 327
Introduction 327
11.1 Temperature and Thermal Energy 327
11.2 Heat, Specific Heat, and Heat Transfer 332
11.3 Phase Change and Latent Heat 340
Key Terms 348
Section Summary 348
Key Equations 349
Chapter Review 349
Test Prep 351

CHAPTER 12
Thermodynamics 355
Introduction 355
12.1 Zeroth Law of Thermodynamics: Thermal Equilibrium 356

12.2 First law of Thermodynamics: Thermal Energy and Work 358
12.3 Second Law of Thermodynamics: Entropy 366
12.4 Applications of Thermodynamics: Heat Engines, Heat Pumps, and Refrigerators 372
Key Terms 378
Section Summary 378
Key Equations 379
Chapter Review 379
Test Prep 383

CHAPTER 13
Waves and Their Properties 389
Introduction 389
13.1 Types of Waves 390
13.2 Wave Properties: Speed, Amplitude, Frequency, and Period 394
13.3 Wave Interaction: Superposition and Interference 400
Key Terms 406
Section Summary 406
Key Equations 407
Chapter Review 407
Test Prep 409

CHAPTER 14
Sound 415
Introduction 415
14.1 Speed of Sound, Frequency, and Wavelength 416
14.2 Sound Intensity and Sound Level 423
14.3 Doppler Effect and Sonic Booms 430
14.4 Sound Interference and Resonance 434
Key Terms 443
Section Summary 443
Key Equations 444
Chapter Review 444
Test Prep 448

CHAPTER 15
Light 455
Introduction 455
15.1 The Electromagnetic Spectrum 455
15.2 The Behavior of Electromagnetic Radiation 463
Key Terms 469
Section Summary 469
Key Equations 469
Chapter Review 469
Test Prep 472

CHAPTER 16
Mirrors and Lenses 477
Introduction 477

16.1 Reflection 478
16.2 Refraction 487
16.3 Lenses 498
Key Terms 513
Section Summary 513
Key Equations 513
Chapter Review 514
Test Prep 517

CHAPTER 17
Diffraction and Interference 523
Introduction 523
17.1 Understanding Diffraction and Interference 523
17.2 Applications of Diffraction, Interference, and Coherence 532
Key Terms 542
Section Summary 542
Key Equations 542
Chapter Review 543
Test Prep 545

CHAPTER 18
Static Electricity 549
Introduction 549
18.1 Electrical Charges, Conservation of Charge, and Transfer of Charge 550
18.2 Coulomb's law 562
18.3 Electric Field 567
18.4 Electric Potential 572
18.5 Capacitors and Dielectrics 580
Key Terms 591
Section Summary 591
Key Equations 592
Chapter Review 592
Test Prep 596

CHAPTER 19
Electrical Circuits 603
Introduction 603
19.1 Ohm's law 604
19.2 Series Circuits 612
19.3 Parallel Circuits 621
19.4 Electric Power 632
Key Terms 638
Section Summary 638
Key Equations 638
Chapter Review 639
Test Prep 643

CHAPTER 20
Magnetism 649
Introduction 649
20.1 Magnetic Fields, Field Lines, and Force 650
20.2 Motors, Generators, and Transformers 665
20.3 Electromagnetic Induction 672
Key Terms 681
Section Summary 681
Key Equations 682
Chapter Review 682
Test Prep 684

CHAPTER 21
The Quantum Nature of Light 691
Introduction 691
21.1 Planck and Quantum Nature of Light 692
21.2 Einstein and the Photoelectric Effect 698
21.3 The Dual Nature of Light 704
Key Terms 711
Section Summary 711
Key Equations 711
Chapter Review 712
Test Prep 715

CHAPTER 22
The Atom 721
Introduction 721
22.1 The Structure of the Atom 721
22.2 Nuclear Forces and Radioactivity 734
22.3 Half Life and Radiometric Dating 743
22.4 Nuclear Fission and Fusion 747
22.5 Medical Applications of Radioactivity: Diagnostic Imaging and Radiation 757
Key Terms 763
Section Summary 763
Key Equations 764
Chapter Review 765
Test Prep 767

CHAPTER 23
Particle Physics 771
Introduction 771
23.1 The Four Fundamental Forces 772
23.2 Quarks 779
23.3 The Unification of Forces 790
Key Terms 796
Section Summary 797
Chapter Review 797
Test Prep 801

Appendix A Reference Tables 807

Index 831

PREFACE

Welcome to *Physics*, an OpenStax resource. This textbook was written to increase student access to high-quality learning materials, maintaining highest standards of academic rigor at little to no cost.

About OpenStax

OpenStax is a nonprofit based at Rice University, and it's our mission to improve student access to education. Our first openly licensed college textbook was published in 2012 and our library has since scaled to over 25 books for college and AP® courses used by hundreds of thousands of students. OpenStax Tutor, our low-cost personalized learning tool, is being used in college courses throughout the country. Through our partnerships with philanthropic foundations and our alliance with other educational resource organizations, OpenStax is breaking down the most common barriers to learning and empowering students and instructors to succeed.

About OpenStax resources

Customization

Physics is licensed under a Creative Commons Attribution 4.0 International (CC BY) license, which means that you can distribute, remix, and build upon the content, as long as you provide attribution to OpenStax and its content contributors.

Because our books are openly licensed, you are free to use the entire book or pick and choose the sections that are most relevant to the needs of your course. Feel free to remix the content by assigning your students certain chapters and sections in your syllabus, in the order that you prefer. You can even provide a direct link in your syllabus to the sections in the web view of your book.

Instructors also have the option of creating a customized version of their OpenStax book. The custom version can be made available to students in low-cost print or digital form through their campus bookstore. Visit your book page on OpenStax.org for more information.

Art Atrribution in *Physics*

In *Physics*, most art contains attribution to its title, creator or rights holder, host platform, and license within the caption. For art that is openly licensed, anyone may reuse the art as long as they provide the same attribution to its original source. Some art has been provided through permissions and should only be used with the attribution or limitations provided in the credit.

Errata

All OpenStax textbooks undergo a rigorous review process. However, like any professional-grade textbook, errors sometimes occur. The good part is, since our books are web-based, we can make updates periodically. If you have a correction to suggest, submit it through our errata reporting tool. We will review your suggestion and make necessary changes.

Format

You can access this textbook for free in web view or PDF through OpenStax.org, and for a low cost in print.

About *Physics*

This instructional material was initially created through a Texas Education Agency (TEA) initiative to provide high-quality open-source instructional materials to districts free of charge. Funds were allocated by the 84th Texas Legislature (2015) for the creation of state-developed, open-source instructional materials with the request that advanced secondary courses supporting the study of science, technology, engineering, and mathematics should be prioritized.

Physics covers the scope and sequence requirements of a typical one-year physics course. The text provides comprehensive coverage of physical concepts, quantitative examples and skills, and interesting applications. High School Physics has been designed to meet and exceed the requirements of the relevant Texas Essential Knowledge and Skills (TEKS (http://ritter.tea.state.tx.us/rules/tac/chapter112/ch112c.html#112.39)), while allowing significant flexibility for instructors.

Qualified and experienced Texas faculty were involved throughout the development process, and the textbooks were reviewed extensively to ensure effectiveness and usability in each course. Reviewers considered each resource's clarity, accuracy, student support, assessment rigor and appropriateness, alignment to TEKS, and overall quality. Their invaluable suggestions provided the basis for continually improved material and helped to certify that the books are ready for use. The writers and reviewers also considered common course issues, effective teaching strategies, and student engagement to provide instructors and students with useful, supportive content and drive effective learning experiences.

Coverage and scope

Physics presents physical laws, research, concepts, and skills in a logical and engaging progression that should be familiar

to most physics faculty. The textbook begins with a general introduction to physics and scientific processes, which is followed by several chapters on motion and Newton's laws. After mechanics, the students will move through thermodynamics, waves and sound, and light and optics. Electricity and magnetism and nuclear physics complete the textbook.

- Chapter 1: What Is Physics?
- Chapter 2: Motion in One Dimension
- Chapter 3: Acceleration
- Chapter 4: Forces and Newton's Laws of Motion
- Chapter 5: Motion in Two Dimensions
- Chapter 6: Circular and Rotational Motion
- Chapter 7: Newton's Law of Gravitation
- Chapter 8: Momentum
- Chapter 9: Work, Energy, and Simple Machines
- Chapter 10: Special Relativity
- Chapter 11: Thermal Energy, Heat, and Work
- Chapter 12: Thermodynamics
- Chapter 13: Waves and Their Properties
- Chapter 14: Sound
- Chapter 15: Light
- Chapter 16: Mirrors and Lenses
- Chapter 17: Diffraction and Interference
- Chapter 18: Static Electricity
- Chapter 19: Electrical Circuits
- Chapter 20: Magnetism
- Chapter 21: The Quantum Nature of Light
- Chapter 22: The Atom
- Chapter 23: Particle Physics

Flexibility

Like any OpenStax content, this textbook can be modified as needed for use by the instructor depending on the needs of the students in the course. Each set of materials created by OpenStax is organized into units and chapters and can be used like a traditional textbook as the entire syllabus for each course. The materials can also be accessed in smaller chunks for more focused use with a single student or an entire class. Instructors are welcome to download and assign the PDF version of the textbook through a learning management system or can use their LMS to link students to specific chapters and sections of the book relevant to the concept being studied. The entire textbook will be available during the fall of 2020 in an editable Google document, and until then instructors are welcome to copy and paste content from the textbook to modify as needed prior to instruction.

Student-centered focus

Physics uses a friendly voice and exciting examples that appeal to a high school audience. The Chapter Openers, for example, include thought-provoking photographs and introductions that connect the content to experiences relevant to student's lives. The writing in our program has been developed with universal design in mind to ensure students of all different backgrounds are reached. Content can be accessed through engaging text, informative visuals, hands-on activities, and online simulations. This diversity of learning media presents a wealth of reinforcement opportunities that allow students to review material in a new and fresh way.

Features

- *Snap Labs*: Give students the opportunity to experience physics through hands-on activities. The labs can be completed quickly and rely primarily on readily available materials so that students can do them at home as they read.
- *Worked Examples*: Promote both analytical and conceptual skills. In each example, the scenario/application is first introduced, followed by a description of the strategy used to solve the problem that emphasizes the concepts involved. These are followed by a fully worked mathematical solution and a discussion of the results.
- *Fun in Physics*: Features physics applications in various entertainment industries.
- *Work in Physics*: Students can explore careers in physics as well as other careers that routinely employ physics.
- *Boundless Physics*: Reveal frontiers in physical knowledge and descriptions of cutting-edge discoveries in physics.
- *Links to Physics*: Highlight connections of physics to other disciplines.
- *Watch Physics*: Support student's understanding of conceptual and computational skills using videos from Khan Academy.
- *Virtual Physics*: Provide inquiry and discovery-based learning by providing a virtual "sandbox" where students can experiment with simulated physics scenarios and equipment using the University of Colorado-developed PhET simulations.
- *Tips for Success*: Offer students advice on how to approach content or problems.

Practice and Assessment

- *Grasp Checks*: Formative assessments that review the comprehension of concepts and skills addressed through reading features, interactive features, and snap labs.
- *Practice Problems*: Challenge students to apply concepts and skills they have seen in a Worked Example to solve a problem.
- *Check Your Understanding*: Conceptual questions that, together with the practice problems, provide formative assessment on key topics in each section.
- *Performance Tasks*: Challenge students to apply the content and skills they have learned to find a solution to a practical situation.
- *Test Prep*: Helps prepare students to successfully respond to the format and rigor of standardized tests. The test prep includes multiple choice, short answer, and extended

response items.

Additional resources

Student and instructor resources

We've compiled additional resources for both students and instructors, including Getting Started Guides, PowerPoint slides, and an instructor answer guide. Instructor resources require a verified instructor account, which you can apply for when you log in or create your account on OpenStax.org. Take advantage of these resources to supplement your OpenStax book.

Partner resources

OpenStax Partners are our allies in the mission to make high-quality learning materials affordable and accessible to students and instructors everywhere. Their tools integrate seamlessly with our OpenStax titles at a low cost. To access the partner resources for your text, visit your book page on OpenStax.org.

About the Authors

Senior Contributing Authors

Paul Peter Urone, California State University, Sacramento
Roger Hinrichs, State University of New York, College at Oswego

Contributing Authors

Fatih Gozuacik (Director of Technology and Education) holds master's degrees in physics and atomic physics from Texas A&M Commerce and Sakarya University in Turkey. In addition to teaching AP® and College Physics, he has been a mentor for the Physics Bowl and the Science Olympiad, and has served as a college counselor for junior and senior students. Fatih was named the 2015 STEM teacher of the year by Educate Texas.

Denise Pattison (East Chambers ISD) was born and raised in Lumberton, Texas. She graduated from Lamar University - Beaumont in 2004 with a degree in science and has taught Physics, Pre AP® Physics, and IPC Physics. She loves teaching because it offers the opportunity to make a difference in a person's life.

Catherine Tabor (Northwest Early High School) holds Bachelors degrees in Mathematics and Physics, a Master's degree in Physics, and is working towards a PhD in Computer Science. She has taught for over twenty years, holding positions at a number of high schools in El Paso and colleges including Evergreen State and UT El Paso. At Northwest Early HS, she teaches Astronomy, AP® Physics, and AP® Computer Science.

Reviewers

- Bryan Callow, Lindenwold High School, Lindenwold, NJ
- John Boehringer, Prosper High School, Prosper, TX
- Wade Green, Stony Point High School, Round Rock, TX
- Susan Vogel, Rockwall-Heath High School, Heath TX
- Stan Hutto, Southwest Christian School, TX
- Richard Lines, Garland High School, Garland, TX
- Penny McCool, San Antonio ISD, Raise Up Texas Coach
- Loren Lykins, Science Department Chairman—Carlisle Independent School District, TX
- Janie Horn, Cleveland High School, Magnolia, TX
- Sean Oshman, Garland Independent School District, TX
- Cort Gillen, Cypress Ranch High School, TX

CHAPTER 1
What is Physics?

Figure 1.1 Galaxies, such as the Andromeda galaxy pictured here, are immense in size. The small blue spots in this photo are also galaxies. The same physical laws apply to objects as large as galaxies or objects as small as atoms. The laws of physics are, therefore, surprisingly few in number. (NASA, JPL-Caltech, P. Barmby, Harvard-Smithsonian Center for Astrophysics).

Chapter Outline

1.1 Physics: Definitions and Applications

1.2 The Scientific Methods

1.3 The Language of Physics: Physical Quantities and Units

INTRODUCTION Take a look at the image above of the Andromeda Galaxy (Figure 1.1), which contains billions of stars. This galaxy is the nearest one to our own galaxy (the Milky Way) but is still a staggering 2.5 million light years from Earth. (A light year is a measurement of the distance light travels in a year.) Yet, the primary force that affects the movement of stars within Andromeda is the same force that we contend with here on Earth—namely, gravity.

You may soon realize that physics plays a much larger role in your life than you thought. This section introduces you to the realm of physics, and discusses applications of physics in other disciplines of study. It also describes the methods by which science is done, and how scientists communicate their results to each other.

1.1 Physics: Definitions and Applications

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Describe the definition, aims, and branches of physics
- Describe and distinguish classical physics from modern physics and describe the importance of relativity, quantum mechanics, and relativistic quantum mechanics in modern physics
- Describe how aspects of physics are used in other sciences (e.g., biology, chemistry, geology, etc.) as well as in everyday technology

Section Key Terms

atom	classical physics	modern physics
physics	quantum mechanics	theory of relativity

What Physics Is

Think about all of the technological devices that you use on a regular basis. Computers, wireless internet, smart phones, tablets, global positioning system (GPS), MP3 players, and satellite radio might come to mind. Next, think about the most exciting modern technologies that you have heard about in the news, such as trains that levitate above their tracks, *invisibility cloaks* that bend light around them, and microscopic robots that fight diseased cells in our bodies. All of these groundbreaking advancements rely on the principles of **physics**.

Physics is a branch of science. The word *science* comes from a Latin word that means *having knowledge*, and refers the knowledge of how the physical world operates, based on objective evidence determined through observation and experimentation. A key requirement of any scientific explanation of a natural phenomenon is that it must be testable; one must be able to devise and conduct an experimental investigation that either supports or refutes the explanation. It is important to note that some questions fall outside the realm of science precisely because they deal with phenomena that are not scientifically testable. This need for objective evidence helps define the investigative process scientists follow, which will be described later in this chapter.

Physics is the science aimed at describing the fundamental aspects of our universe. This includes what things are in it, what properties of those things are noticeable, and what processes those things or their properties undergo. In simpler terms, physics attempts to describe the basic mechanisms that make our universe behave the way it does. For example, consider a smart phone (Figure 1.2). Physics describes how electric current interacts with the various circuits inside the device. This knowledge helps engineers select the appropriate materials and circuit layout when building the smart phone. Next, consider a GPS. Physics describes the relationship between the speed of an object, the distance over which it travels, and the time it takes to travel that distance. When you use a GPS device in a vehicle, it utilizes these physics relationships to determine the travel time from one location to another.

Figure 1.2 Physics describes the way that electric charge flows through the circuits of this device. Engineers use their knowledge of physics to construct a smartphone with features that consumers will enjoy, such as a GPS function. GPS uses physics equations to determine the driving time between two locations on a map. (@gletham GIS, Social, Mobile Tech Images)

As our technology evolved over the centuries, physics expanded into many branches. Ancient peoples could only study things that they could see with the naked eye or otherwise experience without the aid of scientific equipment. This included the study of kinematics, which is the study of moving objects. For example, ancient people often studied the apparent motion of objects in the sky, such as the sun, moon, and stars. This is evident in the construction of prehistoric astronomical observatories, such as Stonehenge in England (shown in Figure 1.3).

Figure 1.3 Stonehenge is a monument located in England that was built between 3000 and 1000 B.C. It functions as an ancient astronomical observatory, with certain rocks in the monument aligning with the position of the sun during the summer and winter solstices. Other rocks align with the rising and setting of the moon during certain days of the year. (Citypeek, Wikimedia Commons)

Ancient people also studied statics and dynamics, which focus on how objects start moving, stop moving, and change speed and direction in response to forces that push or pull on the objects. This early interest in kinematics and dynamics allowed humans to invent simple machines, such as the lever, the pulley, the ramp, and the wheel. These simple machines were gradually

combined and integrated to produce more complicated machines, such as wagons and cranes. Machines allowed humans to gradually do more work more effectively in less time, allowing them to create larger and more complicated buildings and structures, many of which still exist today from ancient times.

As technology advanced, the branches of physics diversified even more. These include branches such as acoustics, the study of sound, and optics, the study of the light. In 1608, the invention of the telescope by a Germany spectacle maker, Hans Lippershey, led to huge breakthroughs in astronomy—the study of objects or phenomena in space. One year later, in 1609, Galileo Galilei began the first studies of the solar system and the universe using a telescope. During the Renaissance era, Isaac Newton used observations made by Galileo to construct his three laws of motion. These laws were the standard for studying kinematics and dynamics even today.

Another major branch of physics is thermodynamics, which includes the study of thermal energy and the transfer of heat. James Prescott Joule, an English physicist, studied the nature of heat and its relationship to work. Joule's work helped lay the foundation for the first of three laws of thermodynamics that describe how energy in our universe is transferred from one object to another or transformed from one form to another. Studies in thermodynamics were motivated by the need to make engines more efficient, keep people safe from the elements, and preserve food.

The 18th and 19th centuries also saw great strides in the study of electricity and magnetism. Electricity involves the study of electric charges and their movements. Magnetism had long ago been noticed as an attractive force between a magnetized object and a metal like iron, or between the opposite poles (North and South) of two magnetized objects. In 1820, Danish physicist Hans Christian Oersted showed that electric currents create magnetic fields. In 1831, English inventor Michael Faraday showed that moving a wire through a magnetic field could induce an electric current. These studies led to the inventions of the electric motor and electric generator, which revolutionized human life by bringing electricity and magnetism into our machines.

The end of the 19th century saw the discovery of radioactive substances by the French scientists Marie and Pierre Curie. Nuclear physics involves studying the nuclei of **atoms**, the source of nuclear radiation. In the 20th century, the study of nuclear physics eventually led to the ability to split the nucleus of an atom, a process called nuclear fission. This process is the basis for nuclear power plants and nuclear weapons. Also, the field of **quantum mechanics**, which involves the mechanics of atoms and molecules, saw great strides during the 20th century as our understanding of atoms and subatomic particles increased (see below).

Early in the 20th century, Albert Einstein revolutionized several branches of physics, especially relativity. Relativity revolutionized our understanding of motion and the universe in general as described further in this chapter. Now, in the 21st century, physicists continue to study these and many other branches of physics.

By studying the most important topics in physics, you will gain analytical abilities that will enable you to apply physics far beyond the scope of what can be included in a single book. These analytical skills will help you to excel academically, and they will also help you to think critically in any career you choose to pursue.

Physics: Past and Present

The word physics is thought to come from the Greek word *phusis*, meaning nature. The study of nature later came to be called *natural philosophy*. From ancient times through the Renaissance, natural philosophy encompassed many fields, including astronomy, biology, chemistry, mathematics, and medicine. Over the last few centuries, the growth of scientific knowledge has resulted in ever-increasing specialization and branching of natural philosophy into separate fields, with physics retaining the most basic facets. Physics, as it developed from the Renaissance to the end of the 19th century, is called **classical physics**. Revolutionary discoveries starting at the beginning of the 20th century transformed physics from classical physics to **modern physics**.

Classical physics is not an exact description of the universe, but it is an excellent approximation under the following conditions: (1) matter must be moving at speeds less than about 1 percent of the speed of light, (2) the objects dealt with must be large enough to be seen with the naked eye, and (3) only weak gravity, such as that generated by Earth, can be involved. Very small objects, such as atoms and molecules, cannot be adequately explained by classical physics. These three conditions apply to almost all of everyday experience. As a result, most aspects of classical physics should make sense on an intuitive level.

Many laws of classical physics have been modified during the 20th century, resulting in revolutionary changes in technology, society, and our view of the universe. As a result, many aspects of modern physics, which occur outside of the range of our everyday experience, may seem bizarre or unbelievable. So why is most of this textbook devoted to classical physics? There are

two main reasons. The first is that knowledge of classical physics is necessary to understand modern physics. The second reason is that classical physics still gives an accurate description of the universe under a wide range of everyday circumstances.

Modern physics includes two revolutionary theories: relativity and quantum mechanics. These theories deal with the very fast and the very small, respectively. The **theory of relativity** was developed by Albert Einstein in 1905. By examining how two observers moving relative to each other would see the same phenomena, Einstein devised radical new ideas about time and space. He came to the startling conclusion that the measured length of an object travelling at high speeds (greater than about one percent of the speed of light) is shorter than the same object measured at rest. Perhaps even more bizarre is the idea the time for the same process to occur is different depending on the motion of the observer. Time passes more slowly for an object travelling at high speeds. A trip to the nearest star system, Alpha Centauri, might take an astronaut 4.5 Earth years if the ship travels near the speed of light. However, because time is slowed at higher speeds, the astronaut would age only 0.5 years during the trip. Einstein's ideas of relativity were accepted after they were confirmed by numerous experiments.

Gravity, the force that holds us to Earth, can also affect time and space. For example, time passes more slowly on Earth's surface than for objects farther from the surface, such as a satellite in orbit. The very accurate clocks on global positioning satellites have to correct for this. They slowly keep getting ahead of clocks at Earth's surface. This is called time dilation, and it occurs because gravity, in essence, slows down time.

Large objects, like Earth, have strong enough gravity to distort space. To visualize this idea, think about a bowling ball placed on a trampoline. The bowling ball depresses or curves the surface of the trampoline. If you rolled a marble across the trampoline, it would follow the surface of the trampoline, roll into the depression caused by the bowling ball, and hit the ball. Similarly, the Earth curves space around it in the shape of a funnel. These curves in space due to the Earth cause objects to be attracted to Earth (i.e., gravity).

Because of the way gravity affects space and time, Einstein stated that gravity affects the space-time continuum, as illustrated in Figure 1.4. This is why time proceeds more slowly at Earth's surface than in orbit. In black holes, whose gravity is hundreds of times that of Earth, time passes so slowly that it would appear to a far-away observer to have stopped!

Figure 1.4 Einstein's theory of relativity describes space and time as an interweaved mesh. Large objects, such as a planet, distort space, causing objects to fall in toward the planet due to the action of gravity. Large objects also distort time, causing time to proceed at a slower rate near the surface of Earth compared with the area outside of the distorted region of space-time.

In summary, relativity says that in describing the universe, it is important to realize that time, space and speed are not absolute. Instead, they can appear different to different observers. Einstein's ability to reason out relativity is even more amazing because we cannot see the effects of relativity in our everyday lives.

Quantum mechanics is the second major theory of modern physics. Quantum mechanics deals with the very small, namely, the subatomic particles that make up atoms. Atoms (Figure 1.5) are the smallest units of elements. However, atoms themselves are constructed of even smaller subatomic particles, such as protons, neutrons and electrons. Quantum mechanics strives to

describe the properties and behavior of these and other subatomic particles. Often, these particles do not behave in the ways expected by classical physics. One reason for this is that they are small enough to travel at great speeds, near the speed of light.

Figure 1.5 Using a scanning tunneling microscope (STM), scientists can see the individual atoms that compose this sheet of gold. (Erwinrossen)

At particle colliders (Figure 1.6), such as the Large Hadron Collider on the France-Swiss border, particle physicists can make subatomic particles travel at very high speeds within a 27 kilometers (17 miles) long superconducting tunnel. They can then study the properties of the particles at high speeds, as well as collide them with each other to see how they exchange energy. This has led to many intriguing discoveries such as the Higgs-Boson particle, which gives matter the property of mass, and antimatter, which causes a huge energy release when it comes in contact with matter.

Figure 1.6 Particle colliders such as the Large Hadron Collider in Switzerland or Fermilab in the United States (pictured here), have long tunnels that allows subatomic particles to be accelerated to near light speed. (Andrius.v)

Physicists are currently trying to unify the two theories of modern physics, relativity and quantum mechanics, into a single, comprehensive theory called relativistic quantum mechanics. Relating the behavior of subatomic particles to gravity, time, and space will allow us to explain how the universe works in a much more comprehensive way.

Application of Physics

You need not be a scientist to use physics. On the contrary, knowledge of physics is useful in everyday situations as well as in nonscientific professions. For example, physics can help you understand why you shouldn't put metal in the microwave (Figure 1.7), why a black car radiator helps remove heat in a car engine, and why a white roof helps keep the inside of a house cool. The operation of a car's ignition system, as well as the transmission of electrical signals through our nervous system, are much easier to understand when you think about them in terms of the basic physics of electricity.

Figure 1.7 Why can't you put metal in the microwave? Microwaves are high-energy radiation that increases the movement of electrons in

metal. These moving electrons can create an electrical current, causing sparking that can lead to a fire. (= MoneyBlogNewz)

Physics is the foundation of many important scientific disciplines. For example, chemistry deals with the interactions of atoms and molecules. Not surprisingly, chemistry is rooted in atomic and molecular physics. Most branches of engineering are also applied physics. In architecture, physics is at the heart of determining structural stability, acoustics, heating, lighting, and cooling for buildings. Parts of geology, the study of nonliving parts of Earth, rely heavily on physics; including radioactive dating, earthquake analysis, and heat transfer across Earth's surface. Indeed, some disciplines, such as biophysics and geophysics, are hybrids of physics and other disciplines.

Physics also describes the chemical processes that power the human body. Physics is involved in medical diagnostics, such as x-rays, magnetic resonance imaging (MRI), and ultrasonic blood flow measurements (Figure 1.8). Medical therapy Physics also has many applications in biology, the study of life. For example, physics describes how cells can protect themselves using their cell walls and cell membranes (Figure 1.9). Medical therapy sometimes directly involves physics, such as in using X-rays to diagnose health conditions. Physics can also explain what we perceive with our senses, such as how the ears detect sound or the eye detects color.

Figure 1.8 Magnetic resonance imaging (MRI) uses electromagnetic waves to yield an image of the brain, which doctors can use to find diseased regions. (Rashmi Chawla, Daniel Smith, and Paul E. Marik)

Figure 1.9 Physics, chemistry, and biology help describe the properties of cell walls in plant cells, such as the onion cells seen here. (Umberto Salvagnin)

BOUNDLESS PHYSICS

The Physics of Landing on a Comet

On November 12, 2014, the European Space Agency's Rosetta spacecraft (shown in Figure 1.10) became the first ever to reach and orbit a comet. Shortly after, Rosetta's rover, Philae, landed on the comet, representing the first time humans have ever landed a space probe on a comet.

Figure 1.10 The Rosetta spacecraft, with its large and revolutionary solar panels, carried the Philae lander to a comet. The lander then detached and landed on the comet's surface. (European Space Agency)

After traveling 6.4 billion kilometers starting from its launch on Earth, Rosetta landed on the comet 67P/Churyumov-Gerasimenko, which is only 4 kilometers wide. Physics was needed to successfully plot the course to reach such a small, distant, and rapidly moving target. Rosetta's path to the comet was not straight forward. The probe first had to travel to Mars so that Mars's gravity could accelerate it and divert it in the exact direction of the comet.

This was not the first time humans used gravity to power our spaceships. Voyager 2, a space probe launched in 1977, used the gravity of Saturn to *slingshot* over to Uranus and Neptune (illustrated in Figure 1.11), providing the first pictures ever taken of these planets. Now, almost 40 years after its launch, Voyager 2 is at the very edge of our solar system and is about to enter interstellar space. Its sister ship, Voyager 1 (illustrated in Figure 1.11), which was also launched in 1977, is already there.

To listen to the sounds of interstellar space or see images that have been transmitted back from the Voyager I or to learn more about the Voyager mission, visit the Voyager's Mission website (https://openstax.org/l/28voyager) .

Figure 1.11 a) Voyager 2, launched in 1977, used the gravity of Saturn to slingshot over to Uranus and Neptune. NASA b) A rendering of Voyager 1, the first space probe to ever leave our solar system and enter interstellar space. NASA

Both Voyagers have electrical power generators based on the decay of radioisotopes. These generators have served them for almost 40 years. Rosetta, on the other hand, is solar-powered. In fact, Rosetta became the first space probe to travel beyond the asteroid belt by relying only on solar cells for power generation.

At 800 million kilometers from the sun, Rosetta receives sunlight that is only 4 percent as strong as on Earth. In addition, it is very cold in space. Therefore, a lot of physics went into developing Rosetta's low-intensity low-temperature solar cells.

In this sense, the Rosetta project nicely shows the huge range of topics encompassed by physics: from modeling the movement of gigantic planets over huge distances within our solar systems, to learning how to generate electric power from low-intensity light. Physics is, by far, the broadest field of science.

GRASP CHECK

What characteristics of the solar system would have to be known or calculated in order to send a probe to a distant planet, such as Jupiter?

a. the effects due to the light from the distant stars
b. the effects due to the air in the solar system
c. the effects due to the gravity from the other planets
d. the effects due to the cosmic microwave background radiation

In summary, physics studies many of the most basic aspects of science. A knowledge of physics is, therefore, necessary to understand all other sciences. This is because physics explains the most basic ways in which our universe works. However, it is not necessary to formally study all applications of physics. A knowledge of the basic laws of physics will be most useful to you, so that you can use them to solve some everyday problems. In this way, the study of physics can improve your problem-solving skills.

Check Your Understanding

1. Which of the following is *not* an essential feature of scientific explanations?
a. They must be subject to testing.
b. They strictly pertain to the physical world.
c. Their validity is judged based on objective observations.
d. Once supported by observation, they can be viewed as a fact.

2. Which of the following does *not* represent a question that can be answered by science?
a. How much energy is released in a given nuclear chain reaction?
b. Can a nuclear chain reaction be controlled?
c. Should uncontrolled nuclear reactions be used for military applications?
d. What is the half-life of a waste product of a nuclear reaction?

3. What are the three conditions under which classical physics provides an excellent description of our universe?
a. 1. Matter is moving at speeds less than about 1 percent of the speed of light
2. Objects dealt with must be large enough to be seen with the naked eye.
3. Strong electromagnetic fields are involved.

b. 1. Matter is moving at speeds less than about 1 percent of the speed of light.
2. Objects dealt with must be large enough to be seen with the naked eye.
3. Only weak gravitational fields are involved.

c. 1. Matter is moving at great speeds, comparable to the speed of light.
2. Objects dealt with are large enough to be seen with the naked eye.
3. Strong gravitational fields are involved.

d. 1. Matter is moving at great speeds, comparable to the speed of light.
2. Objects are just large enough to be visible through the most powerful telescope.
3. Only weak gravitational fields are involved.

4. Why is the Greek word for nature appropriate in describing the field of physics?
a. Physics is a natural science that studies life and living organism on habitable planets like Earth.
b. Physics is a natural science that studies the laws and principles of our universe.
c. Physics is a physical science that studies the composition, structure, and changes of matter in our universe.
d. Physics is a social science that studies the social behavior of living beings on habitable planets like Earth.

5. Which aspect of the universe is studied by quantum mechanics?
a. objects at the galactic level
b. objects at the classical level
c. objects at the subatomic level
d. objects at all levels, from subatomic to galactic

1.2 The Scientific Methods

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Explain how the methods of science are used to make scientific discoveries
- Define a scientific model and describe examples of physical and mathematical models used in physics
- Compare and contrast hypothesis, theory, and law

Section Key Terms

experiment	hypothesis	model	observation	principle
scientific law	scientific methods	theory	universal	

Scientific Methods

Scientists often plan and carry out investigations to answer questions about the universe around us. Such laws are intrinsic to the universe, meaning that humans did not create them and cannot change them. We can only discover and understand them. Their discovery is a very human endeavor, with all the elements of mystery, imagination, struggle, triumph, and disappointment inherent in any creative effort. The cornerstone of discovering natural laws is observation. Science must describe the universe as it is, not as we imagine or wish it to be.

We all are curious to some extent. We look around, make generalizations, and try to understand what we see. For example, we look up and wonder whether one type of cloud signals an oncoming storm. As we become serious about exploring nature, we become more organized and formal in collecting and analyzing data. We attempt greater precision, perform controlled experiments (if we can), and write down ideas about how data may be organized. We then formulate models, theories, and laws based on the data we have collected, and communicate those results with others. This, in a nutshell, describes the **scientific method** that scientists employ to decide scientific issues on the basis of evidence from observation and experiment.

An investigation often begins with a scientist making an **observation**. The scientist observes a pattern or trend within the natural world. Observation may generate questions that the scientist wishes to answer. Next, the scientist may perform some research about the topic and devise a **hypothesis**. A hypothesis is a testable statement that describes how something in the natural world works. In essence, a hypothesis is an educated guess that explains something about an observation.

Scientists may test the hypothesis by performing an **experiment**. During an experiment, the scientist collects data that will help them learn about the phenomenon they are studying. Then the scientists analyze the results of the experiment (that is, the data), often using statistical, mathematical, and/or graphical methods. From the data analysis, they draw conclusions. They may conclude that their experiment either supports or rejects their hypothesis. If the hypothesis is supported, the scientist usually goes on to test another hypothesis related to the first. If their hypothesis is rejected, they will often then test a new and different hypothesis in their effort to learn more about whatever they are studying.

Scientific processes can be applied to many situations. Let's say that you try to turn on your car, but it will not start. You have just made an observation! You ask yourself, "Why won't my car start?" You can now use scientific processes to answer this question. First, you generate a hypothesis such as, "The car won't start because it has no gasoline in the gas tank." To test this hypothesis, you put gasoline in the car and try to start it again. If the car starts, then your hypothesis is supported by the experiment. If the car does not start, then your hypothesis is rejected. You will then need to think up a new hypothesis to test such as, "My car won't start because the fuel pump is broken." Hopefully, your investigations lead you to discover why the car won't start and enable you to fix it.

Modeling

A **model** is a representation of something that is often too difficult (or impossible) to study directly. Models can take the form of physical models, equations, computer programs, or simulations—computer graphics/animations. Models are tools that are especially useful in modern physics because they let us visualize phenomena that we normally cannot observe with our senses, such as very small objects or objects that move at high speeds. For example, we can understand the structure of an atom using models, despite the fact that no one has ever seen an atom with their own eyes. Models are always approximate, so they are simpler to consider than the real situation; the more complete a model is, the more complicated it must be. Models put the

intangible or the extremely complex into human terms that we can visualize, discuss, and hypothesize about.

Scientific models are constructed based on the results of previous experiments. Even still, models often only describe a phenomenon partially or in a few limited situations. Some phenomena are so complex that they may be impossible to model them in their entirety, even using computers. An example is the electron cloud model of the atom in which electrons are moving around the atom's center in distinct clouds (Figure 1.12), that represent the likelihood of finding an electron in different places. This model helps us to visualize the structure of an atom. However, it does not show us exactly where an electron will be within its cloud at any one particular time.

Figure 1.12 The electron cloud model of the atom predicts the geometry and shape of areas where different electrons may be found in an atom. However, it cannot indicate exactly where an electron will be at any one time.

As mentioned previously, physicists use a variety of models including equations, physical models, computer simulations, etc. For example, three-dimensional models are often commonly used in chemistry and physics to model molecules. Properties other than appearance or location are usually modelled using mathematics, where functions are used to show how these properties relate to one another. Processes such as the formation of a star or the planets, can also be modelled using computer simulations. Once a simulation is correctly programmed based on actual experimental data, the simulation can allow us to view processes that happened in the past or happen too quickly or slowly for us to observe directly. In addition, scientists can also run virtual experiments using computer-based models. In a model of planet formation, for example, the scientist could alter the amount or type of rocks present in space and see how it affects planet formation.

Scientists use models and experimental results to construct explanations of observations or design solutions to problems. For example, one way to make a car more fuel efficient is to reduce the friction or drag caused by air flowing around the moving car. This can be done by designing the body shape of the car to be more aerodynamic, such as by using rounded corners instead of sharp ones. Engineers can then construct physical models of the car body, place them in a wind tunnel, and examine the flow of air around the model. This can also be done mathematically in a computer simulation. The air flow pattern can be analyzed for regions smooth air flow and for eddies that indicate drag. The model of the car body may have to be altered slightly to produce the smoothest pattern of air flow (i.e., the least drag). The pattern with the least drag may be the solution to increasing fuel efficiency of the car. This solution might then be incorporated into the car design.

Snap Lab

Using Models and the Scientific Processes

Be sure to secure loose items before opening the window or door.

In this activity, you will learn about scientific models by making a model of how air flows through your classroom or a room in your house.

- One room with at least one window or door that can be opened
- Piece of single-ply tissue paper
  1. Work with a group of four, as directed by your teacher. Close all of the windows and doors in the room you are working in. Your teacher may assign you a specific window or door to study.

2. Before opening any windows or doors, draw a to-scale diagram of your room. First, measure the length and width of your room using the tape measure. Then, transform the measurement using a scale that could fit on your paper, such as 5 centimeters = 1 meter.
3. Your teacher will assign you a specific window or door to study air flow. On your diagram, add arrows showing your hypothesis (before opening any windows or doors) of how air will flow through the room when your assigned window or door is opened. Use pencil so that you can easily make changes to your diagram.
4. On your diagram, mark four locations where you would like to test air flow in your room. To test for airflow, hold a strip of single ply tissue paper between the thumb and index finger. Note the direction that the paper moves when exposed to the airflow. Then, for each location, predict which way the paper will move if your air flow diagram is correct.
5. Now, each member of your group will stand in one of the four selected areas. Each member will test the airflow Agree upon an approximate height at which everyone will hold their papers.
6. When you teacher tells you to, open your assigned window and/or door. Each person should note the direction that their paper points immediately after the window or door was opened. Record your results on your diagram.
7. Did the airflow test data support or refute the hypothetical model of air flow shown in your diagram? Why or why not? Correct your model based on your experimental evidence.
8. With your group, discuss how accurate your model is. What limitations did it have? Write down the limitations that your group agreed upon.

GRASP CHECK

Your diagram is a model, based on experimental evidence, of how air flows through the room. Could you use your model to predict how air would flow through a new window or door placed in a different location in the classroom? Make a new diagram that predicts the room's airflow with the addition of a new window or door. Add a short explanation that describes how.

a. Yes, you could use your model to predict air flow through a new window. The earlier experiment of air flow would help you model the system more accurately.
b. Yes, you could use your model to predict air flow through a new window. The earlier experiment of air flow is not useful for modeling the new system.
c. No, you cannot model a system to predict the air flow through a new window. The earlier experiment of air flow would help you model the system more accurately.
d. No, you cannot model a system to predict the air flow through a new window. The earlier experiment of air flow is not useful for modeling the new system.

Scientific Laws and Theories

A **scientific law** is a description of a pattern in nature that is true in all circumstances that have been studied. That is, physical laws are meant to be **universal**, meaning that they apply throughout the known universe. Laws are often also concise, whereas theories are more complicated. A law can be expressed in the form of a single sentence or mathematical equation. For example, Newton's second law of motion, which relates the motion of an object to the force applied (F), the mass of the object (m), and the object's acceleration (a), is simply stated using the equation

$$F = ma.$$

Scientific ideas and explanations that are true in many, but not all situations in the universe are usually called **principles**. An example is Pascal's principle, which explains properties of liquids, but not solids or gases. However, the distinction between laws and principles is sometimes not carefully made in science.

A **theory** is an explanation for patterns in nature that is supported by much scientific evidence and verified multiple times by multiple researchers. While many people confuse theories with educated guesses or hypotheses, theories have withstood more rigorous testing and verification than hypotheses.

As a closing idea about scientific processes, we want to point out that scientific laws and theories, even those that have been supported by experiments for centuries, can still be changed by new discoveries. This is especially true when new technologies emerge that allow us to observe things that were formerly unobservable. Imagine how viewing previously invisible objects with a

microscope or viewing Earth for the first time from space may have instantly changed our scientific theories and laws! What discoveries still await us in the future? The constant retesting and perfecting of our scientific laws and theories allows our knowledge of nature to progress. For this reason, many scientists are reluctant to say that their studies *prove* anything. By saying *support* instead of *prove*, it keeps the door open for future discoveries, even if they won't occur for centuries or even millennia.

Check Your Understanding

6. Explain why scientists sometimes use a model rather than trying to analyze the behavior of the real system.
   a. Models are simpler to analyze.
   b. Models give more accurate results.
   c. Models provide more reliable predictions.
   d. Models do not require any computer calculations.

7. Describe the difference between a question, generated through observation, and a hypothesis.
   a. They are the same.
   b. A hypothesis has been thoroughly tested and found to be true.
   c. A hypothesis is a tentative assumption based on what is already known.
   d. A hypothesis is a broad explanation firmly supported by evidence.

8. What is a scientific model and how is it useful?
   a. A scientific model is a representation of something that can be easily studied directly. It is useful for studying things that can be easily analyzed by humans.
   b. A scientific model is a representation of something that is often too difficult to study directly. It is useful for studying a complex system or systems that humans cannot observe directly.
   c. A scientific model is a representation of scientific equipment. It is useful for studying working principles of scientific equipment.
   d. A scientific model is a representation of a laboratory where experiments are performed. It is useful for studying requirements needed inside the laboratory.

9. Which of the following statements is correct about the hypothesis?
   a. The hypothesis must be validated by scientific experiments.
   b. The hypothesis must not include any physical quantity.
   c. The hypothesis must be a short and concise statement.
   d. The hypothesis must apply to all the situations in the universe.

10. What is a scientific theory?
   a. A scientific theory is an explanation of natural phenomena that is supported by evidence.
   b. A scientific theory is an explanation of natural phenomena without the support of evidence.
   c. A scientific theory is an educated guess about the natural phenomena occurring in nature.
   d. A scientific theory is an uneducated guess about natural phenomena occurring in nature.

11. Compare and contrast a hypothesis and a scientific theory.
   a. A hypothesis is an explanation of the natural world with experimental support, while a scientific theory is an educated guess about a natural phenomenon.
   b. A hypothesis is an educated guess about natural phenomenon, while a scientific theory is an explanation of natural world with experimental support.
   c. A hypothesis is experimental evidence of a natural phenomenon, while a scientific theory is an explanation of the natural world with experimental support.
   d. A hypothesis is an explanation of the natural world with experimental support, while a scientific theory is experimental evidence of a natural phenomenon.

1.3 The Language of Physics: Physical Quantities and Units

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Associate physical quantities with their International System of Units (SI)and perform conversions among SI units using scientific notation
- Relate measurement uncertainty to significant figures and apply the rules for using significant figures in calculations
- Correctly create, label, and identify relationships in graphs using mathematical relationships (e.g., slope, *y*-intercept, inverse, quadratic and logarithmic)

Section Key Terms

accuracy	ampere	constant	conversion factor	dependent variable
derived units	English units	exponential relationship	fundamental physical units	independent variable
inverse relationship	inversely proportional	kilogram	linear relationship	logarithmic (log) scale
log-log plot	meter	method of adding percents	order of magnitude	precision
quadratic relationship	scientific notation	second	semi-log plot	SI units
significant figures	slope	uncertainty	variable	*y*-intercept

The Role of Units

Physicists, like other scientists, make observations and ask basic questions. For example, how big is an object? How much mass does it have? How far did it travel? To answer these questions, they make measurements with various instruments (e.g., meter stick, balance, stopwatch, etc.).

The measurements of physical quantities are expressed in terms of units, which are standardized values. For example, the length of a race, which is a physical quantity, can be expressed in meters (for sprinters) or kilometers (for long distance runners). Without standardized units, it would be extremely difficult for scientists to express and compare measured values in a meaningful way (Figure 1.13).

Figure 1.13 Distances given in unknown units are maddeningly useless.

All physical quantities in the International System of Units (SI) are expressed in terms of combinations of seven **fundamental**

physical units, which are units for: length, mass, time, electric current, temperature, amount of a substance, and luminous intensity.

SI Units: Fundamental and Derived Units

There are two major systems of units used in the world: **SI units** (acronym for the French Le Système International d'Unités, also known as the metric system), and **English units** (also known as the imperial system). English units were historically used in nations once ruled by the British Empire. Today, the United States is the only country that still uses English units extensively. Virtually every other country in the world now uses the metric system, which is the standard system agreed upon by scientists and mathematicians.

Some physical quantities are more fundamental than others. In physics, there are seven fundamental physical quantities that are measured in base or physical fundamental units: length, mass, time, electric current temperature, amount of substance, and luminous intensity. Units for other physical quantities (such as force, speed, and electric charge) described by mathematically combining these seven base units. In this course, we will mainly use five of these: length, mass, time, electric current and temperature. The units in which they are measured are the meter, kilogram, second, ampere, kelvin, mole, and candela (Table 1.1). All other units are made by mathematically combining the fundamental units. These are called **derived units**.

Quantity	Name	Symbol
Length	Meter	m
Mass	Kilogram	kg
Time	Second	s
Electric current	Ampere	a
Temperature	Kelvin	k
Amount of substance	Mole	mol
Luminous intensity	Candela	cd

Table 1.1 SI Base Units

The Meter

The SI unit for length is the **meter** (m). The definition of the meter has changed over time to become more accurate and precise. The meter was first defined in 1791 as 1/10,000,000 of the distance from the equator to the North Pole. This measurement was improved in 1889 by redefining the meter to be the distance between two engraved lines on a platinum-iridium bar. (The bar is now housed at the International Bureau of Weights and Meaures, near Paris). By 1960, some distances could be measured more precisely by comparing them to wavelengths of light. The meter was redefined as 1,650,763.73 wavelengths of orange light emitted by krypton atoms. In 1983, the meter was given its present definition as the distance light travels in a vacuum in 1/299,792,458 of a second (Figure 1.14).

Figure 1.14 The meter is defined to be the distance light travels in 1/299,792,458 of a second through a vacuum. Distance traveled is speed multiplied by time.

The Kilogram

The SI unit for mass is the **kilogram** (kg). It is defined to be the mass of a platinum-iridium cylinder, housed at the International Bureau of Weights and Measures near Paris. Exact replicas of the standard kilogram cylinder are kept in numerous locations throughout the world, such as the National Institute of Standards and Technology in Gaithersburg, Maryland. The determination of all other masses can be done by comparing them with one of these standard kilograms.

The Second

The SI unit for time, the **second** (s) also has a long history. For many years it was defined as 1/86,400 of an average solar day. However, the average solar day is actually very gradually getting longer due to gradual slowing of Earth's rotation. Accuracy in the fundamental units is essential, since all other measurements are derived from them. Therefore, a new standard was adopted to define the second in terms of a non-varying, or constant, physical phenomenon. One constant phenomenon is the very steady vibration of Cesium atoms, which can be observed and counted. This vibration forms the basis of the cesium atomic clock. In 1967, the second was redefined as the time required for 9,192,631,770 Cesium atom vibrations (Figure 1.15).

Figure 1.15 An atomic clock such as this one uses the vibrations of cesium atoms to keep time to a precision of one microsecond per year. The fundamental unit of time, the second, is based on such clocks. This image is looking down from the top of an atomic clock. (Steve Jurvetson/Flickr)

The Ampere

Electric current is measured in the **ampere** (A), named after Andre Ampere. You have probably heard of amperes, or *amps*, when people discuss electrical currents or electrical devices. Understanding an ampere requires a basic understanding of electricity and magnetism, something that will be explored in depth in later chapters of this book. Basically, two parallel wires with an electric current running through them will produce an attractive force on each other. One ampere is defined as the amount of electric current that will produce an attractive force of 2.7×10^{-7} newton per meter of separation between the two wires (the newton is the derived unit of force).

Kelvins

The SI unit of temperature is the **kelvin** (or kelvins, but not degrees kelvin). This scale is named after physicist William Thomson, Lord Kelvin, who was the first to call for an absolute temperature scale. The Kelvin scale is based on absolute zero. This is the point at which all thermal energy has been removed from all atoms or molecules in a system. This temperature, 0 K, is equal to −273.15 °C and −459.67 °F. Conveniently, the Kelvin scale actually changes in the same way as the Celsius scale. For example, the freezing point (0 °C) and boiling points of water (100 °C) are 100 degrees apart on the Celsius scale. These two temperatures are also 100 kelvins apart (freezing point = 273.15 K; boiling point = 373.15 K).

Metric Prefixes

Physical objects or phenomena may vary widely. For example, the size of objects varies from something very small (like an atom)

to something very large (like a star). Yet the standard metric unit of length is the meter. So, the metric system includes many prefixes that can be attached to a unit. Each prefix is based on factors of 10 (10, 100, 1,000, etc., as well as 0.1, 0.01, 0.001, etc.). Table 1.2 gives the metric prefixes and symbols used to denote the different various factors of 10 in the metric system.

Prefix	Symbol	Value[1]	Example Name	Example Symbol	Example Value	Example Description
exa	E	10^{18}	Exameter	Em	10^{18} m	Distance light travels in a century
peta	P	10^{15}	Petasecond	Ps	10^{15} s	30 million years
tera	T	10^{12}	Terawatt	TW	10^{12} W	Powerful laser output
giga	G	10^{9}	Gigahertz	GHz	10^{9} Hz	A microwave frequency
mega	M	10^{6}	Megacurie	MCi	10^{6} Ci	High radioactivity
kilo	k	10^{3}	Kilometer	km	10^{3} m	About 6/10 mile
hector	h	10^{2}	Hectoliter	hL	10^{2} L	26 gallons
deka	da	10^{1}	Dekagram	dag	10^{1} g	Teaspoon of butter
____	____	10^{0} (=1)				
deci	d	10^{-1}	Deciliter	dL	10^{-1} L	Less than half a soda
centi	c	10^{-2}	Centimeter	Cm	10^{-2} m	Fingertip thickness
milli	m	10^{-3}	Millimeter	Mm	10^{-3} m	Flea at its shoulder
micro	µ	10^{-6}	Micrometer	µm	10^{-6} m	Detail in microscope
nano	n	10^{-9}	Nanogram	Ng	10^{-9} g	Small speck of dust
pico	p	10^{-12}	Picofarad	pF	10^{-12} F	Small capacitor in radio
femto	f	10^{-15}	Femtometer	Fm	10^{-15} m	Size of a proton
atto	a	10^{-18}	Attosecond	as	10^{-18} s	Time light takes to cross an atom

Table 1.2 Metric Prefixes for Powers of 10 and Their Symbols [1]See Appendix A for a discussion of powers of 10. Note—Some examples are approximate.

The metric system is convenient because conversions between metric units can be done simply by moving the decimal place of a number. This is because the metric prefixes are sequential powers of 10. There are 100 centimeters in a meter, 1000 meters in a kilometer, and so on. In nonmetric systems, such as U.S. customary units, the relationships are less simple—there are 12 inches in a foot, 5,280 feet in a mile, 4 quarts in a gallon, and so on. Another advantage of the metric system is that the same unit can be used over extremely large ranges of values simply by switching to the most-appropriate metric prefix. For example, distances in meters are suitable for building construction, but kilometers are used to describe road construction. Therefore, with the metric system, there is no need to invent new units when measuring very small or very large objects—you just have to move the decimal

point (and use the appropriate prefix).

Known Ranges of Length, Mass, and Time

Table 1.3 lists known lengths, masses, and time measurements. You can see that scientists use a range of measurement units. This wide range demonstrates the vastness and complexity of the universe, as well as the breadth of phenomena physicists study. As you examine this table, note how the metric system allows us to discuss and compare an enormous range of phenomena, using one system of measurement (Figure 1.16 and Figure 1.17).

Length (m)	Phenomenon Measured	Mass (Kg)	Phenomenon Measured[1]	Time (s)	Phenomenon Measured[1]
10^{-18}	Present experimental limit to smallest observable detail	10^{-30}	Mass of an electron (9.11×10^{-31} kg)	10^{-23}	Time for light to cross a proton
10^{-15}	Diameter of a proton	10^{-27}	Mass of a hydrogen atom (1.67×10^{-27} kg)	10^{-22}	Mean life of an extremely unstable nucleus
10^{14}	Diameter of a uranium nucleus	10^{-15}	Mass of a bacterium	10^{-15}	Time for one oscillation of a visible light
10^{-10}	Diameter of a hydrogen atom	10^{-5}	Mass of a mosquito	10^{-13}	Time for one vibration of an atom in a solid
10^{-8}	Thickness of membranes in cell of living organism	10^{-2}	Mass of a hummingbird	10^{-8}	Time for one oscillation of an FM radio wave
10^{-6}	Wavelength of visible light	1	Mass of a liter of water (about a quart)	10^{-3}	Duration of a nerve impulse
10^{-3}	Size of a grain of sand	10^{2}	Mass of a person	1	Time for one heartbeat
1	Height of a 4-year-old child	10^{3}	Mass of a car	10^{5}	One day (8.64×10^{4} s)
10^{2}	Length of a football field	10^{8}	Mass of a large ship	10^{7}	One year (3.16×10^{7} s)
10^{4}	Greatest ocean depth	10^{12}	Mass of a large iceberg	10^{9}	About half the life expectancy of a human
10^{7}	Diameter of Earth	10^{15}	Mass of the nucleus of a comet	10^{11}	Recorded history
10^{11}	Distance from Earth to the sun	10^{23}	Mass of the moon (7.35×10^{22} kg)	10^{17}	Age of Earth
10^{16}	Distance traveled by light in 1 year (a light year)	10^{25}	Mass of Earth (5.97×10^{24} kg)	10^{18}	Age of the universe
10^{21}	Diameter of the Milky Way Galaxy	10^{30}	Mass of the Sun (1.99×10^{24} kg)		

Table 1.3 Approximate Values of Length, Mass, and Time [1] More precise values are in parentheses.

Length (m)	Phenomenon Measured	Mass (Kg)	Phenomenon Measured[1]	Time (s)	Phenomenon Measured[1]
10^{22}	Distance from Earth to the nearest large galaxy (Andromeda)	10^{42}	Mass of the Milky Way galaxy (current upper limit)		
10^{26}	Distance from Earth to the edges of the known universe	10^{53}	Mass of the known universe (current upper limit)		

Table 1.3 Approximate Values of Length, Mass, and Time [1] More precise values are in parentheses.

Figure 1.16 Tiny phytoplankton float among crystals of ice in the Antarctic Sea. They range from a few micrometers to as much as 2 millimeters in length. (Prof. Gordon T. Taylor, Stony Brook University; NOAA Corps Collections)

Figure 1.17 Galaxies collide 2.4 billion light years away from Earth. The tremendous range of observable phenomena in nature challenges the imagination. (NASA/CXC/UVic./A. Mahdavi et al. Optical/lensing: CFHT/UVic./H. Hoekstra et al.)

Using Scientific Notation with Physical Measurements

Scientific notation is a way of writing numbers that are too large or small to be conveniently written as a decimal. For example, consider the number 840,000,000,000,000. It's a rather large number to write out. The scientific notation for this number is 8.40×10^{14}. Scientific notation follows this general format

$$x \times 10^{y}.$$

In this format x is the value of the measurement with all placeholder zeros removed. In the example above, x is 8.4. The x is multiplied by a factor, 10^{y}, which indicates the number of placeholder zeros in the measurement. Placeholder zeros are those at the end of a number that is 10 or greater, and at the beginning of a decimal number that is less than 1. In the example above, the factor is 10^{14}. This tells you that you should move the decimal point 14 positions to the right, filling in placeholder zeros as you go. In this case, moving the decimal point 14 places creates only 13 placeholder zeros, indicating that the actual measurement value is 840,000,000,000,000.

Numbers that are fractions can be indicated by scientific notation as well. Consider the number 0.0000045. Its scientific notation is 4.5×10^{-6}. Its scientific notation has the same format

$$x \times 10^{y}.$$

Here, x is 4.5. However, the value of y in the 10^{y} factor is negative, which indicates that the measurement is a fraction of 1. Therefore, we move the decimal place to the left, for a negative y. In our example of 4.5×10^{-6}, the decimal point would be moved to the left six times to yield the original number, which would be 0.0000045.

The term **order of magnitude** refers to the power of 10 when numbers are expressed in scientific notation. Quantities that have the same power of 10 when expressed in scientific notation, or come close to it, are said to be of the same order of magnitude. For example, the number 800 can be written as 8×10^{2}, and the number 450 can be written as 4.5×10^{2}. Both numbers have the same value for y. Therefore, 800 and 450 are of the same order of magnitude. Similarly, 101 and 99 would be regarded as the same order of magnitude, 10^{2}. Order of magnitude can be thought of as a ballpark estimate for the scale of a value. The diameter of an atom is on the order of 10^{-9} m, while the diameter of the sun is on the order of 10^{9} m. These two values are 18 orders of magnitude apart.

Scientists make frequent use of scientific notation because of the vast range of physical measurements possible in the universe, such as the distance from Earth to the moon (Figure 1.18), or to the nearest star.

Figure 1.18 The distance from Earth to the moon may seem immense, but it is just a tiny fraction of the distance from Earth to our closest neighboring star. (NASA)

Unit Conversion and Dimensional Analysis

It is often necessary to convert from one type of unit to another. For example, if you are reading a European cookbook in the United States, some quantities may be expressed in liters and you need to convert them to cups. A Canadian tourist driving through the United States might want to convert miles to kilometers, to have a sense of how far away his next destination is. A doctor in the United States might convert a patient's weight in pounds to kilograms.

Let's consider a simple example of how to convert units within the metric system. How can we want to convert 1 hour to seconds?

Next, we need to determine a **conversion factor** relating meters to kilometers. A **conversion factor** is a ratio expressing how many of one unit are equal to another unit. A conversion factor is simply a fraction which equals 1. You can multiply any number by 1 and get the same value. When you multiply a number by a conversion factor, you are simply multiplying it by one. For example, the following are conversion factors: (1 foot)/(12 inches) = 1 to convert inches to feet, (1 meter)/(100 centimeters) = 1 to convert centimeters to meters, (1 minute)/(60 seconds) = 1 to convert seconds to minutes. In this case, we know that there are 1,000 meters in 1 kilometer.

Now we can set up our unit conversion. We will write the units that we have and then multiply them by the conversion factor (1 km/1,000m) = 1, so we are simply multiplying 80m by 1:

$$1\ \cancel{h} \times \frac{60\ \cancel{min}}{1\ \cancel{h}} \times \frac{60\ \text{s}}{1\ \cancel{min}} = 3600\ \text{s} = 3.6 \times 10^{2}\ \text{s} \tag{1.1}$$

When there is a unit in the original number, and a unit in the denominator (bottom) of the conversion factor, the units cancel. In this case, hours and minutes cancel and the value in seconds remains.

You can use this method to convert between any types of unit, including between the U.S. customary system and metric system. Notice also that, although you can multiply and divide units algebraically, you cannot add or subtract different units. An expression like *10 km + 5 kg* makes no sense. Even adding two lengths in different units, such as *10 km + 20 m* does not make sense. You express both lengths in the same unit. See Appendix C for a more complete list of conversion factors.

WORKED EXAMPLE

Unit Conversions: A Short Drive Home

Suppose that you drive the 10.0 km from your university to home in 20.0 min. Calculate your average speed (a) in kilometers per hour (km/h) and (b) in meters per second (m/s). (Note—Average speed is distance traveled divided by time of travel.)

Strategy

First we calculate the average speed using the given units. Then we can get the average speed into the desired units by picking the correct conversion factor and multiplying by it. The correct conversion factor is the one that cancels the unwanted unit and leaves the desired unit in its place.

Solution for (a)

1. Calculate average speed. Average speed is distance traveled divided by time of travel. (Take this definition as a given for now—average speed and other motion concepts will be covered in a later module.) In equation form,

$$\text{average speed} = \frac{\text{distance}}{\text{time}}.$$

2. Substitute the given values for distance and time.

$$\text{average speed} = \frac{10.0\ \text{km}}{20.0\ \text{min}} = 0.500\ \frac{\text{km}}{\text{min}}$$

3. Convert km/min to km/h: multiply by the conversion factor that will cancel minutes and leave hours. That conversion factor is $60\ \text{min}/1\text{h}$. Thus,

$$\text{average speed} = 0.500\ \frac{\text{km}}{\text{min}} \times \frac{60\ \text{min}}{1\ \text{h}} = 30.0\ \frac{\text{km}}{\text{h}}.$$

Discussion for (a)

To check your answer, consider the following:

1. Be sure that you have properly cancelled the units in the unit conversion. If you have written the unit conversion factor upside down, the units will not cancel properly in the equation. If you accidentally get the ratio upside down, then the units will not cancel; rather, they will give you the wrong units as follows

$$\frac{\text{km}}{\text{min}} \times \frac{1\ \text{hr}}{60\ \text{min}} = \frac{1}{60}\frac{\text{km}\cdot\text{h}}{\text{min}^2},$$

   which are obviously not the desired units of km/h.

2. Check that the units of the final answer are the desired units. The problem asked us to solve for average speed in units of km/h and we have indeed obtained these units.
3. Check the significant figures. Because each of the values given in the problem has three significant figures, the answer should also have three significant figures. The answer 30.0 km/h does indeed have three significant figures, so this is appropriate. Note that the significant figures in the conversion factor are not relevant because an hour is *defined* to be 60 min, so the precision of the conversion factor is perfect.
4. Next, check whether the answer is reasonable. Let us consider some information from the problem—if you travel 10 km in a third of an hour (20 min), you would travel three times that far in an hour. The answer does seem reasonable.

Solution (b)

There are several ways to convert the average speed into meters per second.

1. Start with the answer to (a) and convert km/h to m/s. Two conversion factors are needed—one to convert hours to seconds, and another to convert kilometers to meters.
2. Multiplying by these yields

$$\text{Averagespeed} = 30.0\frac{\text{km}}{\text{h}} \times \frac{1\ \text{h}}{3{,}600\ \text{s}} \times \frac{1{,}000\ \text{m}}{1\ \text{km}}$$

$$\text{Averagespeed} = 8.33\,\frac{\text{m}}{\text{s}}$$

Discussion for (b)

If we had started with 0.500 km/min, we would have needed different conversion factors, but the answer would have been the same: 8.33 m/s.

You may have noted that the answers in the worked example just covered were given to three digits. Why? When do you need to be concerned about the number of digits in something you calculate? Why not write down all the digits your calculator produces?

WORKED EXAMPLE

Using Physics to Evaluate Promotional Materials

A commemorative coin that is 2" in diameter is advertised to be plated with 15 mg of gold. If the density of gold is 19.3 g/cc, and the amount of gold around the edge of the coin can be ignored, what is the thickness of the gold on the top and bottom faces of the coin?

Strategy

To solve this problem, the volume of the gold needs to be determined using the gold's mass and density. Half of that volume is distributed on each face of the coin, and, for each face, the gold can be represented as a cylinder that is 2" in diameter with a height equal to the thickness. Use the volume formula for a cylinder to determine the thickness.

Solution

The mass of the gold is given by the formula $m = \rho V = 15 \times 10^{-3}\ \text{g}$, where $\rho = 19.3\ \text{g/cc}$ and V is the volume. Solving for the volume gives $V = \frac{m}{\rho} = \frac{15 \times 10^{-3}\ \text{g}}{19.3\ \text{g/cc}} \cong 7.8 \times 10^{-4}\ \text{cc}$.

If t is the thickness, the volume corresponding to half the gold is $\frac{1}{2}\left(7.8 \times 10^{-4}\right) = \pi r^2 t = \pi(2.54)^2\text{t}$, where the 1" radius has been converted to cm. Solving for the thickness gives $t = \frac{\left(3.9 \times 10^{-4}\right)}{\pi(2.54)^2} \cong 1.9 \times 10^{-5}\ \text{cm} = 0.00019\ \text{mm}$.

Discussion

The amount of gold used is stated to be 15 mg, which is equivalent to a thickness of about 0.00019 mm. The mass figure may make the amount of gold sound larger, both because the number is much bigger (15 versus 0.00019), and because people may have a more intuitive feel for how much a millimeter is than for how much a milligram is. A simple analysis of this sort can clarify the significance of claims made by advertisers.

Accuracy, Precision and Significant Figures

Science is based on experimentation that requires good measurements. The validity of a measurement can be described in terms of its accuracy and its precision (see Figure 1.19 and Figure 1.20). **Accuracy** is how close a measurement is to the correct value for that measurement. For example, let us say that you are measuring the length of standard piece of printer paper. The packaging in which you purchased the paper states that it is 11 inches long, and suppose this stated value is correct. You measure the length of the paper three times and obtain the following measurements: 11.1 inches, 11.2 inches, and 10.9 inches. These measurements are quite accurate because they are very close to the correct value of 11.0 inches. In contrast, if you had obtained a measurement of 12 inches, your measurement would not be very accurate. This is why measuring instruments are calibrated based on a known measurement. If the instrument consistently returns the correct value of the known measurement, it is safe for use in finding unknown values.

Figure 1.19 A double-pan mechanical balance is used to compare different masses. Usually an object with unknown mass is placed in one pan and objects of known mass are placed in the other pan. When the bar that connects the two pans is horizontal, then the masses in both pans are equal. The known masses are typically metal cylinders of standard mass such as 1 gram, 10 grams, and 100 grams. (Serge Melki)

Figure 1.20 Whereas a mechanical balance may only read the mass of an object to the nearest tenth of a gram, some digital scales can measure the mass of an object up to the nearest thousandth of a gram. As in other measuring devices, the precision of a scale is limited to the last measured figures. This is the hundredths place in the scale pictured here. (Splarka, Wikimedia Commons)

Precision states how well repeated measurements of something generate the same or similar results. Therefore, the precision of measurements refers to how close together the measurements are when you measure the same thing several times. One way to analyze the precision of measurements would be to determine the range, or difference between the lowest and the highest measured values. In the case of the printer paper measurements, the lowest value was 10.9 inches and the highest value was 11.2 inches. Thus, the measured values deviated from each other by, at most, 0.3 inches. These measurements were reasonably precise because they varied by only a fraction of an inch. However, if the measured values had been 10.9 inches, 11.1 inches, and 11.9 inches, then the measurements would not be very precise because there is a lot of variation from one measurement to another.

The measurements in the paper example are both accurate and precise, but in some cases, measurements are accurate but not precise, or they are precise but not accurate. Let us consider a GPS system that is attempting to locate the position of a restaurant in a city. Think of the restaurant location as existing at the center of a bull's-eye target. Then think of each GPS attempt to locate the restaurant as a black dot on the bull's eye.

In Figure 1.21, you can see that the GPS measurements are spread far apart from each other, but they are all relatively close to the actual location of the restaurant at the center of the target. This indicates a low precision, high accuracy measuring system. However, in Figure 1.22, the GPS measurements are concentrated quite closely to one another, but they are far away from the target location. This indicates a high precision, low accuracy measuring system. Finally, in Figure 1.23, the GPS is both precise and accurate, allowing the restaurant to be located.

Figure 1.21 A GPS system attempts to locate a restaurant at the center of the bull's-eye. The black dots represent each attempt to pinpoint the location of the restaurant. The dots are spread out quite far apart from one another, indicating low precision, but they are each rather close to the actual location of the restaurant, indicating high accuracy. (Dark Evil)

Figure 1.22 In this figure, the dots are concentrated close to one another, indicating high precision, but they are rather far away from the actual location of the restaurant, indicating low accuracy. (Dark Evil)

Figure 1.23 In this figure, the dots are concentrated close to one another, indicating high precision, but they are rather far away from the actual location of the restaurant, indicating low accuracy. (Dark Evil)

Uncertainty

The accuracy and precision of a measuring system determine the **uncertainty** of its measurements. Uncertainty is a way to describe how much your measured value deviates from the actual value that the object has. If your measurements are not very accurate or precise, then the uncertainty of your values will be very high. In more general terms, uncertainty can be thought of as a disclaimer for your measured values. For example, if someone asked you to provide the mileage on your car, you might say that it is 45,000 miles, plus or minus 500 miles. The plus or minus amount is the uncertainty in your value. That is, you are indicating that the actual mileage of your car might be as low as 44,500 miles or as high as 45,500 miles, or anywhere in between. All measurements contain some amount of uncertainty. In our example of measuring the length of the paper, we might say that the length of the paper is 11 inches plus or minus 0.2 inches or 11.0 ± 0.2 inches. The uncertainty in a

measurement, A, is often denoted as δA ("delta A"),

The factors contributing to uncertainty in a measurement include the following:

1. Limitations of the measuring device
2. The skill of the person making the measurement
3. Irregularities in the object being measured
4. Any other factors that affect the outcome (highly dependent on the situation)

In the printer paper example uncertainty could be caused by: the fact that the smallest division on the ruler is 0.1 inches, the person using the ruler has bad eyesight, or uncertainty caused by the paper cutting machine (e.g., one side of the paper is slightly longer than the other.) It is good practice to carefully consider all possible sources of uncertainty in a measurement and reduce or eliminate them,

Percent Uncertainty

One method of expressing uncertainty is as a percent of the measured value. If a measurement, A, is expressed with uncertainty, δA, the percent uncertainty is

$$\%\ \text{uncertainty} = \frac{\delta \mathrm{A}}{\mathrm{A}} \times 100\%. \tag{1.2}$$

WORKED EXAMPLE

Calculating Percent Uncertainty: A Bag of Apples

A grocery store sells 5-lb bags of apples. You purchase four bags over the course of a month and weigh the apples each time. You obtain the following measurements:

- Week 1 weight: $4.\ 8\text{lb}$
- Week 2 weight: $5.3\ \text{lb}$
- Week 3 weight: $4.\ 9\text{lb}$
- Week 4 weight: $5.4\ \text{lb}$

You determine that the weight of the 5 lb bag has an uncertainty of ±0.4 lb. What is the percent uncertainty of the bag's weight?

Strategy

First, observe that the expected value of the bag's weight, A, is 5 lb. The uncertainty in this value, δA, is 0.4 lb. We can use the following equation to determine the percent uncertainty of the weight

$$\%\ \text{uncertainty} = \frac{\delta \mathrm{A}}{\mathrm{A}} \times 100\%.$$

Solution

Plug the known values into the equation

$$\%\ \text{uncertainty} = \frac{0.4\ \text{lb}}{5\ \text{lb}} \times 100\% = 8\%.$$

Discussion

We can conclude that the weight of the apple bag is 5 lb ± 8 percent. Consider how this percent uncertainty would change if the bag of apples were half as heavy, but the uncertainty in the weight remained the same. Hint for future calculations: when calculating percent uncertainty, always remember that you must multiply the fraction by 100 percent. If you do not do this, you will have a decimal quantity, not a percent value.

Uncertainty in Calculations

There is an uncertainty in anything calculated from measured quantities. For example, the area of a floor calculated from measurements of its length and width has an uncertainty because the both the length and width have uncertainties. How big is the uncertainty in something you calculate by multiplication or division? If the measurements in the calculation have small uncertainties (a few percent or less), then the **method of adding percents** can be used. This method says that the percent uncertainty in a quantity calculated by multiplication or division is the sum of the percent uncertainties in the items used to make the calculation. For example, if a floor has a length of 4.00 m and a width of 3.00 m, with uncertainties of 2 percent and 1

percent, respectively, then the area of the floor is 12.0 m^2 and has an uncertainty of 3 percent (expressed as an area this is 0.36 m^2, which we round to 0.4 m^2 since the area of the floor is given to a tenth of a square meter).

For a quick demonstration of the accuracy, precision, and uncertainty of measurements based upon the units of measurement, try this simulation (http://openstax.org/l/28precision) . You will have the opportunity to measure the length and weight of a desk, using milli- versus centi- units. Which do you think will provide greater accuracy, precision and uncertainty when measuring the desk and the notepad in the simulation? Consider how the nature of the hypothesis or research question might influence how precise of a measuring tool you need to collect data.

Precision of Measuring Tools and Significant Figures

An important factor in the accuracy and precision of measurements is the precision of the measuring tool. In general, a precise measuring tool is one that can measure values in very small increments. For example, consider measuring the thickness of a coin. A standard ruler can measure thickness to the nearest millimeter, while a micrometer can measure the thickness to the nearest 0.005 millimeter. The micrometer is a more precise measuring tool because it can measure extremely small differences in thickness. The more precise the measuring tool, the more precise and accurate the measurements can be.

When we express measured values, we can only list as many digits as we initially measured with our measuring tool (such as the rulers shown in Figure 1.24). For example, if you use a standard ruler to measure the length of a stick, you may measure it with a decimeter ruler as 3.6 cm. You could not express this value as 3.65 cm because your measuring tool was not precise enough to measure a hundredth of a centimeter. It should be noted that the last digit in a measured value has been estimated in some way by the person performing the measurement. For example, the person measuring the length of a stick with a ruler notices that the stick length seems to be somewhere in between 36 mm and 37 mm. He or she must estimate the value of the last digit. The rule is that the last digit written down in a measurement is the first digit with some uncertainty. For example, the last measured value 36.5 mm has three digits, or three significant figures. The number of **significant figures** in a measurement indicates the precision of the measuring tool. The more precise a measuring tool is, the greater the number of significant figures it can report.

Figure 1.24 Three metric rulers are shown. The first ruler is in decimeters and can measure point three decimeters. The second ruler is in centimeters long and can measure three point six centimeters. The last ruler is in millimeters and can measure thirty-six point five millimeters.

Zeros

Special consideration is given to zeros when counting significant figures. For example, the zeros in 0.053 are not significant because they are only placeholders that locate the decimal point. There are two significant figures in 0.053—the 5 and the 3. However, if the zero occurs between other significant figures, the zeros are significant. For example, both zeros in 10.053 are significant, as these zeros were actually measured. Therefore, the 10.053 placeholder has five significant figures. The zeros in 1300 may or may not be significant, depending on the style of writing numbers. They could mean the number is known to the last zero, or the zeros could be placeholders. So 1300 could have two, three, or four significant figures. To avoid this ambiguity,

write 1300 in scientific notation as 1.3×10^3. Only significant figures are given in the x factor for a number in scientific notation (in the form $x \times 10^y$). Therefore, we know that 1 and 3 are the only significant digits in this number. In summary, zeros are significant except when they serve only as placeholders. Table 1.4 provides examples of the number of significant figures in various numbers.

Number	Significant Figures	Rationale
1.657	4	There are no zeros and all non-zero numbers are always significant.
0.4578	4	The first zero is only a placeholder for the decimal point.
0.000458	3	The first four zeros are placeholders needed to report the data to the ten-thoudsandths place.
2000.56	6	The three zeros are significant here because they occur between other significant figures.
45,600	3	With no underlines or scientific notation, we assume that the last two zeros are placeholders and are not significant.
15895000	7	The two underlined zeros are significant, while the last zero is not, as it is not underlined.
5.457×10^{13}	4	In scientific notation, all numbers reported in front of the multiplication sign are significant
6.520×10^{-23}	4	In scientific notation, all numbers reported in front of the multiplication sign are significant, including zeros.

Table 1.4

Significant Figures in Calculations

When combining measurements with different degrees of accuracy and precision, the number of significant digits in the final answer can be no greater than the number of significant digits in the least precise measured value. There are two different rules, one for multiplication and division and another rule for addition and subtraction, as discussed below.

1. **For multiplication and division:** The answer should have the same number of significant figures as the starting value with the fewest significant figures. For example, the area of a circle can be calculated from its radius using $A = \pi r^2$. Let us see how many significant figures the area will have if the radius has only two significant figures, for example, $r = 2.0$ m. Then, using a calculator that keeps eight significant figures, you would get

$$A = \pi r^2 = (3.1415927...) \times (2.0 \text{ m})^2 = 4.5238934 \text{ m}^2.$$

But because the radius has only two significant figures, the area calculated is meaningful only to two significant figures or

$$A = 4.5 \text{ m}^2$$

even though the value of π is meaningful to at least eight digits.

2. **For addition and subtraction**: The answer should have the same number places (e.g. tens place, ones place, tenths place, etc.) as the least-precise starting value. Suppose that you buy 7.56 kg of potatoes in a grocery store as measured with a scale having a precision of 0.01 kg. Then you drop off 6.052 kg of potatoes at your laboratory as measured by a scale with a precision of 0.001 kg. Finally, you go home and add 13.7 kg of potatoes as measured by a bathroom scale with a precision of 0.1 kg. How many kilograms of potatoes do you now have, and how many significant figures are appropriate in the answer? The mass is found by simple addition and subtraction:

$$\begin{array}{rl} 7.56 & \text{kg} \\ -6.052 & \text{kg} \\ +13.7 & \text{kg} \\ \hline 15.208 & \text{kg} \end{array}$$

The least precise measurement is 13.7 kg. This measurement is expressed to the 0.1 decimal place, so our final answer must also be expressed to the 0.1 decimal place. Thus, the answer should be rounded to the tenths place, giving 15.2 kg. The same is true for non-decimal numbers. For example,

$$6527.23 + 2 = 6528.23 = 6528\ .$$

We cannot report the decimal places in the answer because *2* has no decimal places that would be significant. Therefore, we can only report to the ones place.

It is a good idea to keep extra significant figures while calculating, and to round off to the correct number of significant figures only in the final answers. The reason is that small errors from rounding while calculating can sometimes produce significant errors in the final answer. As an example, try calculating $5{,}098 - (5.000) \times (1{,}010)$ to obtain a final answer to only two significant figures. Keeping all significant during the calculation gives 48. Rounding to two significant figures in the middle of the calculation changes it to $5{,}100 - (5.000) \times (1{,}000) = 100$, which is way off. You would similarly avoid rounding in the middle of the calculation in counting and in doing accounting, where many small numbers need to be added and subtracted accurately to give possibly much larger final numbers.

Significant Figures in this Text

In this textbook, most numbers are assumed to have three significant figures. Furthermore, consistent numbers of significant figures are used in all worked examples. You will note that an answer given to three digits is based on input good to at least three digits. If the input has fewer significant figures, the answer will also have fewer significant figures. Care is also taken that the number of significant figures is reasonable for the situation posed. In some topics, such as optics, more than three significant figures will be used. Finally, if a number is exact, such as the *2* in the formula, $c = 2\pi r$, it does not affect the number of significant figures in a calculation.

WORKED EXAMPLE

Approximating Vast Numbers: a Trillion Dollars

The U.S. federal deficit in the 2008 fiscal year was a little greater than $10 trillion. Most of us do not have any concept of how much even one trillion actually is. Suppose that you were given a trillion dollars in $100 bills. If you made 100-bill stacks, like that shown in Figure 1.25, and used them to evenly cover a football field (between the end zones), make an approximation of how high the money pile would become. (We will use feet/inches rather than meters here because football fields are measured in yards.) One of your friends says 3 in., while another says 10 ft. What do you think?

Figure 1.25 A bank stack contains one hundred $100 bills, and is worth $10,000. How many bank stacks make up a trillion dollars? (Andrew Magill)

Strategy

When you imagine the situation, you probably envision thousands of small stacks of 100 wrapped $100 bills, such as you might see in movies or at a bank. Since this is an easy-to-approximate quantity, let us start there. We can find the volume of a stack of 100 bills, find out how many stacks make up one trillion dollars, and then set this volume equal to the area of the football field multiplied by the unknown height.

Solution

1. Calculate the volume of a stack of 100 bills. The dimensions of a single bill are approximately 3 in. by 6 in. A stack of 100 of these is about 0.5 in. thick. So the total volume of a stack of 100 bills is

$$\text{volume of stack} = \text{length} \times \text{width} \times \text{height},$$
$$\text{volume of stack} = 6 \text{ in.} \times 3 \text{ in.} \times 0.5 \text{ in.},$$
$$\text{volume of stack} = 9 \text{ in.}^3.$$

2. Calculate the number of stacks. Note that a trillion dollars is equal to $\$1 \times 10^{12}$, and a stack of one-hundred $\$100$ bills is equal to $\$10,000$, or $\$1 \times 10^4$. The number of stacks you will have is

$$\$1 \times 10^{12} \text{ (a trillion dollars)} / \$1 \times 10^4 \text{ per stack} = 1 \times 10^8 \text{ stacks.} \quad 1.3$$

3. Calculate the area of a football field in square inches. The area of a football field is $100 \text{ yd} \times 50\text{yd}$, which gives $5,000 \text{ yd}^2$. Because we are working in inches, we need to convert square yards to square inches

$$\text{Area} = 5{,}000 \text{ yd}^2 \times \frac{3 \text{ ft}}{1\text{yd}} \times \frac{3 \text{ ft}}{1\text{yd}} \times \frac{12 \text{ in.}}{1 \text{ foot}} \times \frac{12 \text{ in.}}{1 \text{ foot}} = 6{,}480{,}000 \text{ in.}^2,$$
$$\text{Area} \approx 6 \times 10^6 \text{ in.}^2.$$

This conversion gives us $6 \times 10^6 \text{ in.}^2$ for the area of the field. (Note that we are using only one significant figure in these calculations.)

4. Calculate the total volume of the bills. The volume of all the $100-bill stacks is
$9 \text{ in.}^3 / \text{stack} \times 10^8 \text{ stacks} = 9 \times 10^8 \text{ in.}^3$
5. Calculate the height. To determine the height of the bills, use the following equation

$$\begin{aligned}
\text{volume of bills} &= \text{area of field} \times \text{height of money} \\
\text{Height of money} &= \frac{\text{volume of bills}}{\text{area of field}} \\
\text{Height of money} &= \frac{9 \times 10^8 \text{ in.}^3}{6 \times 10^6 \text{ in.}^2} = 1.33 \times 10^2 \text{ in.} \\
\text{Height of money} &= 1 \times 10^2 \text{ in.} = 100 \text{ in.}
\end{aligned}$$

The height of the money will be about 100 in. high. Converting this value to feet gives

$$100 \text{ in.} \times \frac{1\text{ft}}{12 \text{ in.}} = 8.33 \text{ ft} \approx 8 \text{ ft.}$$

Discussion

The final approximate value is much higher than the early estimate of 3 in., but the other early estimate of 10 ft (120 in.) was roughly correct. How did the approximation measure up to your first guess? What can this exercise tell you in terms of rough *guesstimates* versus carefully calculated approximations?

In the example above, the final approximate value is much higher than the first friend's early estimate of 3 in. However, the other friend's early estimate of 10 ft. (120 in.) was roughly correct. How did the approximation measure up to your first guess? What can this exercise suggest about the value of rough *guesstimates* versus carefully calculated approximations?

Graphing in Physics

Most results in science are presented in scientific journal articles using graphs. Graphs present data in a way that is easy to visualize for humans in general, especially someone unfamiliar with what is being studied. They are also useful for presenting large amounts of data or data with complicated trends in an easily-readable way.

One commonly-used graph in physics and other sciences is the line graph, probably because it is the best graph for showing how one quantity changes in response to the other. Let's build a line graph based on the data in Table 1.5, which shows the measured distance that a train travels from its station versus time. Our two **variables**, or things that change along the graph, are time in minutes, and distance from the station, in kilometers. Remember that measured data may not have perfect accuracy.

Time (min)	Distance from Station (km)
0	0
10	24
20	36
30	60
40	84
50	97
60	116
70	140

Table 1.5

1. Draw the two axes. The horizontal axis, or *x*-axis, shows the **independent variable**, which is the variable that is controlled or manipulated. The vertical axis, or *y*-axis, shows the **dependent variable**, the non-manipulated variable that changes with (or is dependent on) the value of the independent variable. In the data above, time is the independent variable and should be plotted on the *x*-axis. Distance from the station is the dependent variable and should be plotted on the *y*-axis.

2. Label each axes on the graph with the name of each variable, followed by the symbol for its units in parentheses. Be sure to leave room so that you can number each axis. In this example, use *Time (min)* as the label for the *x*-axis.
3. Next, you must determine the best scale to use for numbering each axis. Because the time values on the *x*-axis are taken every 10 minutes, we could easily number the *x*-axis from 0 to 70 minutes with a tick mark every 10 minutes. Likewise, the *y*-axis scale should start low enough and continue high enough to include all of the *distance from station* values. A scale from 0 km to 160 km should suffice, perhaps with a tick mark every 10 km.

   In general, you want to pick a scale for both axes that 1) shows all of your data, and 2) makes it easy to identify trends in your data. If you make your scale too large, it will be harder to see how your data change. Likewise, the smaller and more fine you make your scale, the more space you will need to make the graph. The number of significant figures in the axis values should be coarser than the number of significant figures in the measurements.

4. Now that your axes are ready, you can begin plotting your data. For the first data point, count along the *x*-axis until you find the 10 min tick mark. Then, count up from that point to the 10 km tick mark on the *y*-axis, and approximate where 22 km is along the *y*-axis. Place a dot at this location. Repeat for the other six data points (Figure 1.26).

Figure 1.26 The graph of the train's distance from the station versus time from the exercise above.

5. Add a title to the top of the graph to state what the graph is describing, such as the *y*-axis parameter vs. the *x*-axis parameter. In the graph shown here, the title is *train motion*. It could also be titled distance of the train from the station vs. time.
6. Finally, with data points now on the graph, you should draw a trend line (Figure 1.27). The trend line represents the dependence you think the graph represents, so that the person who looks at your graph can see how close it is to the real data. In the present case, since the data points look like they ought to fall on a straight line, you would draw a straight line as the trend line. Draw it to come closest to all the points. Real data may have some inaccuracies, and the plotted points may not all fall on the trend line. In some cases, none of the data points fall exactly on the trend line.

Figure 1.27 The completed graph with the trend line included.

Analyzing a Graph Using Its Equation

One way to get a quick snapshot of a dataset is to look at the equation of its trend line. If the graph produces a straight line, the equation of the trend line takes the form

$$y = mx + b.$$

The b in the equation is the y-intercept while the m in the equation is the **slope**. The y-intercept tells you at what y value the line intersects the y-axis. In the case of the graph above, the y-intercept occurs at 0, at the very beginning of the graph. The y-intercept, therefore, lets you know immediately where on the y-axis the plot line begins.

The m in the equation is the slope. This value describes how much the line on the graph moves up or down on the y-axis along the line's length. The slope is found using the following equation

$$m = \frac{Y_2 - Y_1}{X_2 - X_1}.$$

In order to solve this equation, you need to pick two points on the line (preferably far apart on the line so the slope you calculate describes the line accurately). The quantities Y_2 and Y_1 represent the y-values from the two points on the line (not data points) that you picked, while X_2 and X_1 represent the two x-values of the those points.

What can the slope value tell you about the graph? The slope of a perfectly horizontal line will equal zero, while the slope of a perfectly vertical line will be undefined because you cannot divide by zero. A positive slope indicates that the line moves up the y-axis as the x-value increases while a negative slope means that the line moves down the y-axis. The more negative or positive the slope is, the steeper the line moves up or down, respectively. The slope of our graph in Figure 1.26 is calculated below based on the two endpoints of the line

$$\begin{aligned} m &= \frac{Y_2 - Y_1}{X_2 - X_1} \\ m &= \frac{(80\text{ km}) - (20\text{ km})}{(40\text{ min}) - (10\text{ min})} \\ m &= \frac{60\text{ km}}{30\text{ min}} \\ m &= 2.0\text{ km/min}. \end{aligned}$$

Equation of line: $y = (2.0\ \text{km/min})\,x + 0$

Because the x axis is time in minutes, we would actually be more likely to use the time t as the independent (x-axis) variable and write the equation as

$$y = (2.0 \text{ km/min})\, t + 0. \tag{1.4}$$

The formula $y = mx + b$ only applies to **linear relationships**, or ones that produce a straight line. Another common type of line in physics is the **quadratic relationship**, which occurs when one of the variables is squared. One quadratic relationship in physics is the relation between the speed of an object its centripetal acceleration, which is used to determine the force needed to keep an object moving in a circle. Another common relationship in physics is the **inverse relationship**, in which one variable decreases whenever the other variable increases. An example in physics is Coulomb's law. As the distance between two charged objects increases, the electrical force between the two charged objects decreases. **Inverse proportionality**, such the relation between x and y in the equation

$$y = k/x,$$

for some number k, is one particular kind of inverse relationship. A third commonly-seen relationship is the **exponential relationship**, in which a change in the independent variable produces a proportional change in the dependent variable. As the value of the dependent variable gets larger, its rate of growth also increases. For example, bacteria often reproduce at an exponential rate when grown under ideal conditions. As each generation passes, there are more and more bacteria to reproduce. As a result, the growth rate of the bacterial population increases every generation (Figure 1.28).

Figure 1.28 Examples of (a) linear, (b) quadratic, (c) inverse, and (d) exponential relationship graphs.

Using Logarithmic Scales in Graphing

Sometimes a variable can have a very large range of values. This presents a problem when you're trying to figure out the best scale to use for your graph's axes. One option is to use a **logarithmic (log) scale**. In a logarithmic scale, the value each mark labels

is the previous mark's value multiplied by some constant. For a log base 10 scale, each mark labels a value that is 10 times the value of the mark before it. Therefore, a base 10 logarithmic scale would be numbered: 0, 10, 100, 1,000, etc. You can see how the logarithmic scale covers a much larger range of values than the corresponding linear scale, in which the marks would label the values 0, 10, 20, 30, and so on.

If you use a logarithmic scale on one axis of the graph and a linear scale on the other axis, you are using a **semi-log plot**. The Richter scale, which measures the strength of earthquakes, uses a semi-log plot. The degree of ground movement is plotted on a logarithmic scale against the assigned intensity level of the earthquake, which ranges linearly from 1-10 (Figure 1.29 **(a)**).

If a graph has both axes in a logarithmic scale, then it is referred to as a **log-log plot**. The relationship between the wavelength and frequency of electromagnetic radiation such as light is usually shown as a log-log plot (Figure 1.29 **(b)**). Log-log plots are also commonly used to describe exponential functions, such as radioactive decay.

Figure 1.29 (a) The Richter scale uses a log base 10 scale on its *y-axis* (microns of amplified maximum ground motion). (b) The relationship between the frequency and wavelength of electromagnetic radiation can be plotted as a straight line if a log-log plot is used.

Virtual Physics

Graphing Lines

In this simulation you will examine how changing the slope and *y*-intercept of an equation changes the appearance of a plotted line. Select slope-intercept form and drag the blue circles along the line to change the line's characteristics. Then, play the line game and see if you can determine the slope or *y*-intercept of a given line.

Click to view content (https://phet.colorado.edu/sims/html/graphing-lines/latest/graphing-lines_en.html)

GRASP CHECK

How would the following changes affect a line that is neither horizontal nor vertical and has a positive slope?

1. increase the slope but keeping the *y*-intercept constant
2. increase the *y*-intercept but keeping the slope constant
   a. Increasing the slope will cause the line to rotate clockwise around the *y*-intercept. Increasing the *y*-intercept will cause the line to move vertically up on the graph without changing the line's slope.
   b. Increasing the slope will cause the line to rotate counter-clockwise around the *y*-intercept. Increasing the *y*-intercept will cause the line to move vertically up on the graph without changing the line's slope.
   c. Increasing the slope will cause the line to rotate clockwise around the *y*-intercept. Increasing the *y*-intercept will cause the line to move horizontally right on the graph without changing the line's slope.

d. Increasing the slope will cause the line to rotate counter-clockwise around the *y*-intercept. Increasing the *y*-intercept will cause the line to move horizontally right on the graph without changing the line's slope.

Check Your Understanding

12. Identify some advantages of metric units.
a. Conversion between units is easier in metric units.
b. Comparison of physical quantities is easy in metric units.
c. Metric units are more modern than English units.
d. Metric units are based on powers of 2.

13. The length of an American football field is $100\ \text{yd}$, excluding the end zones. How long is the field in meters? Round to the nearest $0.1\ \text{m}$.
a. $10.2\ \text{m}$
b. $91.4\ \text{m}$
c. $109.4\ \text{m}$
d. $328.1\ \text{m}$

14. The speed limit on some interstate highways is roughly $100\ \text{km/h}$. How many miles per hour is this if $1.0\ \text{mile}$ is about $1.609\ \text{km}$?
a. 0.1 mi/h
b. 27.8 mi/h
c. 62 mi/h
d. 160 mi/h

15. Briefly describe the target patterns for accuracy and precision and explain the differences between the two.
a. Precision states how much repeated measurements generate the same or closely similar results, while accuracy states how close a measurement is to the true value of the measurement.
b. Precision states how close a measurement is to the true value of the measurement, while accuracy states how much repeated measurements generate the same or closely similar result.
c. Precision and accuracy are the same thing. They state how much repeated measurements generate the same or closely similar results.
d. Precision and accuracy are the same thing. They state how close a measurement is to the true value of the measurement.

KEY TERMS

accuracy how close a measurement is to the correct value for that measurement

ampere the SI unit for electrical current

atom smallest and most basic units of matter

classical physics physics, as it developed from the Renaissance to the end of the nineteenth century

constant a quantity that does not change

conversion factor a ratio expressing how many of one unit are equal to another unit

dependent variable the vertical, or *y*-axis, variable, which changes with (or is dependent on) the value of the independent variable

derived units units that are derived by combining the fundamental physical units

English units (also known as the customary or imperial system) system of measurement used in the United States; includes units of measurement such as feet, gallons, degrees Fahrenheit, and pounds

experiment process involved with testing a hypothesis

exponential relationship relation between variables in which a constant change in the independent variable is accompanied by change in the dependent variable that is proportional to the value it already had

fundamental physical units the seven fundamental physical units in the SI system of units are length, mass, time, electric current, temperature, amount of a substance, and luminous intensity

hypothesis testable statement that describes how something in the natural world works

independent variable the horizontal, or *x*-axis, variable, which is not influence by the second variable on the graph, the dependent variable

inverse proportionality a relation between two variables expressible by an equation of the form $y = k/x$ where k stays constant when x and y change; the special form of inverse relationship that satisfies this equation

inverse relationship any relation between variables where one variable decreases as the other variable increases

kilogram the SI unit for mass, abbreviated (kg)

linear relationships relation between variables that produce a straight line when graphed

log-log plot a plot that uses a logarithmic scale in both axes

logarithmic scale a graphing scale in which each tick on an axis is the previous tick multiplied by some value

meter the SI unit for length, abbreviated (m)

method of adding percents calculating the percent uncertainty of a quantity in multiplication or division by adding the percent uncertainties in the quantities being added or divided

model system that is analogous to the real system of interest in essential ways but more easily analyzed

modern physics physics as developed from the twentieth century to the present, involving the theories of relativity and quantum mechanics

observation step where a scientist observes a pattern or trend within the natural world

order of magnitude the size of a quantity in terms of its power of 10 when expressed in scientific notation

physics science aimed at describing the fundamental aspects of our universe—energy, matter, space, motion, and time

precision how well repeated measurements generate the same or closely similar results

principle description of nature that is true in many, but not all situations

quadratic relationship relation between variables that can be expressed in the form $y = ax^2 + bx + c$, which produces a curved line when graphed

quantum mechanics major theory of modern physics which describes the properties and nature of atoms and their subatomic particles

science the study or knowledge of how the physical world operates, based on objective evidence determined through observation and experimentation

scientific law pattern in nature that is true in all circumstances studied thus far

scientific methods techniques and processes used in the constructing and testing of scientific hypotheses, laws, and theories, and in deciding issues on the basis of experiment and observation

scientific notation way of writing numbers that are too large or small to be conveniently written in simple decimal form; the measurement is multiplied by a power of 10, which indicates the number of placeholder zeros in the measurement

second the SI unit for time, abbreviated (s)

semi-log plot A plot that uses a logarithmic scale on one axis of the graph and a linear scale on the other axis.

SI units International System of Units (SI); the international system of units that scientists in most countries have agreed to use; includes units such as meters, liters, and grams; also known as the metric system

significant figures when writing a number, the digits, or number of digits, that express the precision of a measuring tool used to measure the number

slope the ratio of the change of a graph on the *y* axis to the change along the *x*-axis, the value of m in the equation of a line, $y = mx + b$

theory explanation of patterns in nature that is supported by much scientific evidence and verified multiple times by various groups of researchers

theory of relativity theory constructed by Albert Einstein which describes how space, time and energy are different

for different observers in relative motion

uncertainty a quantitative measure of how much measured values deviate from a standard or expected value

universal applies throughout the known universe

***y*-intercept** the point where a plot line intersects the *y*-axis

SECTION SUMMARY

1.1 Physics: Definitions and Applications

- Physics is the most fundamental of the sciences, concerning itself with energy, matter, space and time, and their interactions.
- Modern physics involves the theory of relativity, which describes how time, space and gravity are not constant in our universe can be different for different observers, and quantum mechanics, which describes the behavior of subatomic particles.
- Physics is the basis for all other sciences, such as chemistry, biology and geology, because physics describes the fundamental way in which the universe functions.

1.2 The Scientific Methods

- Science seeks to discover and describe the underlying order and simplicity in nature.
- The processes of science include observation, hypothesis, experiment, and conclusion.
- Theories are scientific explanations that are supported by a large body experimental results.
- Scientific laws are concise descriptions of the universe that are universally true.

1.3 The Language of Physics: Physical Quantities and Units

- Physical quantities are a characteristic or property of an object that can be measured or calculated from other measurements.
- The four fundamental units we will use in this textbook are the meter (for length), the kilogram (for mass), the second (for time), and the ampere (for electric current). These units are part of the metric system, which uses powers of 10 to relate quantities over the vast ranges encountered in nature.
- Unit conversions involve changing a value expressed in one type of unit to another type of unit. This is done by using conversion factors, which are ratios relating equal quantities of different units.
- Accuracy of a measured value refers to how close a measurement is to the correct value. The uncertainty in a measurement is an estimate of the amount by which the measurement result may differ from this value.
- Precision of measured values refers to how close the agreement is between repeated measurements.
- Significant figures express the precision of a measuring tool.
- When multiplying or dividing measured values, the final answer can contain only as many significant figures as the least precise value.
- When adding or subtracting measured values, the final answer cannot contain more decimal places than the least precise value.

KEY EQUATIONS

1.3 The Language of Physics: Physical Quantities and Units

slope intercept form	$y = mx + b$
quadratic formula	$y = ax^2 + bx + c$
positive exponential formula	$y = a^x$
negative exponential formula	$y = a^{-x}$

CHAPTER REVIEW

Concept Items

1.1 Physics: Definitions and Applications

1. Which statement best compares and contrasts the aims and topics of natural philosophy had versus physics?
   a. Natural philosophy included all aspects of nature including physics.
   b. Natural philosophy included all aspects of nature excluding physics.
   c. Natural philosophy and physics are different.
   d. Natural philosophy and physics are essentially the

same thing.

2. Which of the following is not an underlying assumption essential to scientific understanding?
   a. Characteristics of the physical universe can be perceived and objectively measured by human beings.
   b. Explanations of natural phenomena can be established with absolute certainty.
   c. Fundamental physical processes dictate how characteristics of the physical universe evolve.
   d. The fundamental processes of nature operate the same way everywhere and at all times.

3. Which of the following questions regarding a strain of genetically modified rice is not one that can be answered by science?
   a. How does the yield of the genetically modified rice compare with that of existing rice?
   b. Is the genetically modified rice more resistant to infestation than existing rice?
   c. How does the nutritional value of the genetically modified rice compare to that of existing rice?
   d. Should the genetically modified rice be grown commercially and sold in the marketplace?

4. What conditions imply that we can use classical physics without considering special relativity or quantum mechanics?
   a. 1. matter is moving at speeds of less than roughly 1 percent the speed of light,
      2. objects are large enough to be seen with the naked eye, and
      3. there is the involvement of a strong gravitational field.
   b. 1. matter is moving at speeds greater than roughly 1 percent the speed of light,
      2. objects are large enough to be seen with the naked eye, and
      3. there is the involvement of a strong gravitational field.
   c. 1. matter is moving at speeds of less than roughly 1 percent the speed of light,
      2. objects are too small to be seen with the naked eye, and
      3. there is the involvement of only a weak gravitational field.
   d. 1. matter is moving at speeds of less than roughly 1 percent the speed of light,
      2. objects are large enough to be seen with the naked eye, and
      3. there is the involvement of a weak gravitational field.

5. How could physics be useful in weather prediction?
   a. Physics helps in predicting how burning fossil fuel releases pollutants.
   b. Physics helps in predicting dynamics and movement of weather phenomena.
   c. Physics helps in predicting the motion of tectonic plates.
   d. Physics helps in predicting how the flowing water affects Earth's surface.

6. How do physical therapists use physics while on the job? Explain.
   a. Physical therapists do not require knowledge of physics because their job is mainly therapy and not physics.
   b. Physical therapists do not require knowledge of physics because their job is more social in nature and unscientific.
   c. Physical therapists require knowledge of physics know about muscle contraction and release of energy.
   d. Physical therapists require knowledge of physics to know about chemical reactions inside the body and make decisions accordingly.

7. What is meant when a physical law is said to be universal?
   a. The law can explain everything in the universe.
   b. The law is applicable to all physical phenomena.
   c. The law applies everywhere in the universe.
   d. The law is the most basic one and all laws are derived from it.

8. What subfield of physics could describe small objects traveling at high speeds or experiencing a strong gravitational field?
   a. general theory of relativity
   b. classical physics
   c. quantum relativity
   d. special theory of relativity

9. Why is Einstein's theory of relativity considered part of modern physics, as opposed to classical physics?
   a. Because it was considered less outstanding than the classics of physics, such as classical mechanics.
   b. Because it was popular physics enjoyed by average people today, instead of physics studied by the elite.
   c. Because the theory deals with very slow-moving objects and weak gravitational fields.
   d. Because it was among the new 19th-century discoveries that changed physics.

1.2 The Scientific Methods

10. Describe the difference between an observation and a hypothesis.

a. An observation is seeing what happens; a hypothesis is a testable, educated guess.
b. An observation is a hypothesis that has been confirmed.
c. Hypotheses and observations are independent of each other.
d. Hypotheses are conclusions based on some observations.

11. Describe how modeling is useful in studying the structure of the atom.
a. Modeling replaces the real system by something similar but easier to examine.
b. Modeling replaces the real system by something more interesting to examine.
c. Modeling replaces the real system by something with more realistic properties.
d. Modeling includes more details than are present in the real system.

12. How strongly is a hypothesis supported by evidence compared to a theory?
a. A theory is supported by little evidence, if any, at first, while a hypothesis is supported by a large amount of available evidence.
b. A hypothesis is supported by little evidence, if any, at first. A theory is supported by a large amount of available evidence.
c. A hypothesis is supported by little evidence, if any, at first. A theory does not need any experiments in support.
d. A theory is supported by little evidence, if any, at first. A hypothesis does not need any experiments in support.

1.3 The Language of Physics: Physical Quantities and Units

13. Which of the following does not contribute to the uncertainty?
a. the limitations of the measuring device
b. the skill of the person making the measurement
c. the regularities in the object being measured
d. other factors that affect the outcome (depending on the situation)

14. How does the independent variable in a graph differ from the dependent variable?
a. The dependent variable varies linearly with the independent variable.
b. The dependent variable depends on the scale of the axis chosen while independent variable does not.
c. The independent variable is directly manipulated or controlled by the person doing the experiment, while dependent variable is the one that changes as a result.
d. The dependent and independent variables are fixed by a convention and hence they are the same.

15. What could you conclude about these two lines?

1. Line A has a slope of -4.7
2. Line B has a slope of 12.0

a. Line A is a decreasing line while line B is an increasing line, with line A being much steeper than line B.
b. Line A is a decreasing line while line B is an increasing line, with line B being much steeper than line A.
c. Line B is a decreasing line while line A is an increasing line, with line A being much steeper than line B.
d. Line B is a decreasing line while line A is an increasing line, with line B being much steeper than line A.

16. Velocity, or speed, is measured using the following formula: $v = \frac{d}{t}$, where v is velocity, d is the distance travelled, and t is the time the object took to travel the distance. If the velocity-time data are plotted on a graph, which variable will be on which axis? Why?
a. Time would be on the x-axis and velocity on the y-axis, because time is an independent variable and velocity is a dependent variable.
b. Velocity would be on the x-axis and time on the y-axis, because time is the independent variable and velocity is the dependent variable.
c. Time would be on the x-axis and velocity on the y-axis, because time is a dependent variable and velocity is a independent variable.
d. Velocity would be on x-axis and time on the y-axis, because time is a dependent variable and velocity is a independent variable.

17. The uncertainty of a triple-beam balance is $0.05\ \text{g}$. What is the percent uncertainty in a measurement of $0.445\ \text{kg}$?
a. 0.011%
b. 0.11%
c. 1.1%
d. 11%

18. What is the definition of uncertainty?
a. Uncertainty is the number of assumptions made prior to the measurement of a physical quantity.
b. Uncertainty is a measure of error in a measurement due to the use of a non-calibrated instrument.
c. Uncertainty is a measure of deviation of the measured value from the standard value.
d. Uncertainty is a measure of error in measurement

due to external factors like air friction and temperature.

Critical Thinking Items

1.1 Physics: Definitions and Applications

19. In what sense does Einstein's theory of relativity illustrate that physics describes fundamental aspects of our universe?
   a. It describes how speed affects different observers' measurements of time and space.
   b. It describes how different parts of the universe are far apart and do not affect each other.
   c. It describes how people think of other people's views from their own frame of reference.
   d. It describes how a frame of reference is necessary to describe position or motion.

20. Can classical physics be used to accurately describe a satellite moving at a speed of 7500 m/s? Explain why or why not.
   a. No, because the satellite is moving at a speed much smaller than the speed of the light and is not in a strong gravitational field.
   b. No, because the satellite is moving at a speed much smaller than the speed of the light and is in a strong gravitational field.
   c. Yes, because the satellite is moving at a speed much smaller than the speed of the light and it is not in a strong gravitational field.
   d. Yes, because the satellite is moving at a speed much smaller than the speed of the light and is in a strong gravitational field.

21. What would be some ways in which physics was involved in building the features of the room you are in right now?
   a. Physics is involved in structural strength, dimensions, etc., of the room.
   b. Physics is involved in the air composition inside the room.
   c. Physics is involved in the desk arrangement inside the room.
   d. Physics is involved in the behavior of living beings inside the room.

22. What theory of modern physics describes the interrelationships between space, time, speed, and gravity?
   a. atomic theory
   b. nuclear physics
   c. quantum mechanics
   d. general relativity

23. According to Einstein's theory of relativity, how could you effectively travel many years into Earth's future, but not age very much yourself?
   a. by traveling at a speed equal to the speed of light
   b. by traveling at a speed faster than the speed of light
   c. by traveling at a speed much slower than the speed of light
   d. by traveling at a speed slightly slower than the speed of light

1.2 The Scientific Methods

24. You notice that the water level flowing in a stream near your house increases when it rains and the water turns brown. Which of these are the best hypothesis to explain why the water turns brown. Assume you have all of the means to test the contents of the stream water.
   a. The water in the stream turns brown because molecular forces between water molecules are stronger than mud molecules
   b. The water in the stream turns brown because of the breakage of a weak chemical bond with the hydrogen atom in the water molecule.
   c. The water in the stream turns brown because it picks up dirt from the bank as the water level increases when it rains.
   d. The water in the stream turns brown because the density of the water increases with increase in water level.

25. Light travels as waves at an approximate speed of 300,000,000 m/s (186,000 mi/s). Designers of devices that use mirrors and lenses model the traveling light by straight lines, or light rays. Describe why it would be useful to model the light as rays of light instead of describing them accurately as electromagnetic waves.
   a. A model can be constructed in such a way that the speed of light decreases.
   b. Studying a model makes it easier to analyze the path that the light follows.
   c. Studying a model will help us to visualize why light travels at such great speed.
   d. Modeling cannot be used to study traveling light as our eyes cannot track the motion of light.

26. A friend says that he doesn't trust scientific explanations because they are just theories, which are basically educated guesses. What could you say to convince him that scientific theories are different from the everyday use of the word theory?
   a. A theory is a scientific explanation that has been repeatedly tested and supported by many experiments.
   b. A theory is a hypothesis that has been tested and

supported by some experiments.
c. A theory is a set of educated guesses, but at least one of the guesses remain true in each experiment.
d. A theory is a set of scientific explanations that has at least one experiment in support of it.

27. Give an example of a hypothesis that cannot be tested experimentally.
a. The structure of any part of the broccoli is similar to the whole structure of the broccoli.
b. Ghosts are the souls of people who have died.
c. The average speed of air molecules increases with temperature.
d. A vegetarian is less likely to be affected by night blindness.

28. Would it be possible to scientifically prove that a supreme being exists or not? Briefly explain your answer.
a. It can be proved scientifically because it is a testable hypothesis.
b. It cannot be proved scientifically because it is not a testable hypothesis.
c. It can be proved scientifically because it is not a testable hypothesis.
d. It cannot be proved scientifically because it is a testable hypothesis.

1.3 The Language of Physics: Physical Quantities and Units

29. A marathon runner completes a $42.188\ \text{km}$ course in $2\ \text{h}$, $30\ \text{min}$, and $12\ \text{s}$. There is an uncertainty of $25\ \text{m}$ in the distance traveled and an uncertainty of $1\ \text{s}$ in the elapsed time.

1. Calculate the percent uncertainty in the distance.
2. Calculate the uncertainty in the elapsed time.
3. What is the average speed in meters per second?
4. What is the uncertainty in the average speed?

a. $0.059\ \%, 0.01\ \%, 0.468\text{m/s}, 0.0003\text{m/s}$
b. $0.059\ \%, 0.01\ \%, 0.468\ \text{m/s}, 0.07\ \text{m/s}$
c. $0.59\ \%, 8.33\ \%, 4.681\ \text{m/s}, 0.003\ \text{m/s}$
d. $0.059\ \%, 0.01\ \%, 4.681\ \text{m/s}, 0.003\ \text{m/s}$

30. A car engine moves a piston with a circular cross section of $7.500 \pm 0.002\ \text{cm}$ diameter a distance of $3.250 \pm 0.001\ \text{cm}$ to compress the gas in the cylinder. By what amount did the gas decrease in volume in cubic centimeters? Find the uncertainty in this volume.
a. $143.6 \pm 0.002\ \text{cm}^3$
b. $143.6 \pm 0.003\ \text{cm}^3$
c. $143.6 \pm 0.005\ \text{cm}^3$
d. $143.6 \pm 0.1\ \text{cm}^3$

31. What would be the slope for a line passing through the two points below?

Point 1: (1, 0.1) Point 2: (7, 26.8)

a. 2.4
b. 4.5
c. 6.2
d. 6.8

32. The sides of a small rectangular box are measured $1.80\ \text{cm}$ and $2.05\ \text{cm}$ long and $3.1\ \text{cm}$ high. Calculate its volume and uncertainty in cubic centimeters. Assume the measuring device is accurate to $\pm 0.05\ \text{cm}$.
a. $11.4 \pm 0.1\ \text{cm}^3$
b. $11.4 \pm 0.6\ \text{cm}^3$
c. $11.4 \pm 0.8\ \text{cm}^3$
d. $11.4 \pm 0.10\ \text{cm}^3$

33. Calculate the approximate number of atoms in a bacterium. Assume that the average mass of an atom in the bacterium is ten times the mass of a hydrogen atom. (Hint—The mass of a hydrogen atom is on the order of 10^{-27} kg and the mass of a bacterium is on the order of 10^{-15} kg .)
a. 10^{10} atoms
b. 10^{11} atoms
c. 10^{12} atoms
d. 10^{13} atoms

Problems

1.3 The Language of Physics: Physical Quantities and Units

34. A commemorative coin that sells for $40 is advertised to be plated with 15 mg of gold. Suppose gold is worth about $1,300 per ounce. Which of the following best represents the value of the gold in the coin?
a. $0.33
b. $0.69
c. $3.30
d. $6.90

35. If a marathon runner runs $9.5\ \text{miles}$ in one direction, $8.89\ \text{miles}$ in another direction and $2.333\ \text{miles}$ in a third direction, how much distance did the runner run? Be sure to report your answer using the proper number of significant figures.
a. 20
b. 20.7
c. 20.72

d. 20.732

36. The speed limit on some interstate highways is roughly 80 km/h. What is this in meters per second? How many miles per hour is this?
a. 62 m/s, 27.8 mi/h
b. 22.2 m/s, 49.7 mi/h
c. 62 m/s, 2.78 mi/h
d. 2.78 m/s, 62 mi/h

37. The length and width of a rectangular room are measured to be 3.955 ± 0.005 m by 3.050 ± 0.005 m. Calculate the area of the room and its uncertainty in square meters.
a. $12.06 \pm 0.29\ \text{m}^2$
b. $12.06 \pm 0.01\ \text{m}^2$
c. $12.06 \pm 0.25\ \text{m}^2$
d. $12.06 \pm 0.04\ \text{m}^2$

Performance Task

1.3 The Language of Physics: Physical Quantities and Units

38. a. Create a new system of units to describe something that interests you. Your unit should be described using at least two subunits. For example, you can decide to measure the quality of songs using a new unit called *song awesomeness*. Song awesomeness is measured by: the number of songs downloaded and the number of times the song was used in movies.
b. Create an equation that shows how to calculate your unit. Then, using your equation, create a sample dataset that you could graph. Are your two subunits related linearly, quadratically, or inversely?

TEST PREP

Multiple Choice

1.1 Physics: Definitions and Applications

39. Modern physics could best be described as the combination of which theories?
a. quantum mechanics and Einstein's theory of relativity
b. quantum mechanics and classical physics
c. Newton's laws of motion and classical physics
d. Newton's laws of motion and Einstein's theory of relativity

40. Which of the following could be studied accurately using classical physics?
a. the strength of gravity within a black hole
b. the motion of a plane through the sky
c. the collisions of subatomic particles
d. the effect of gravity on the passage of time

41. Which of the following best describes why knowledge of physics is necessary to understand all other sciences?
a. Physics explains how energy passes from one object to another.
b. Physics explains how gravity works.
c. Physics explains the motion of objects that can be seen with the naked eye.
d. Physics explains the fundamental aspects of the universe.

42. What does radiation therapy, used to treat cancer patients, have to do with physics?
a. Understanding how cells reproduce is mainly about physics.
b. Predictions of the side effects from the radiation therapy are based on physics.
c. The devices used for generating some kinds of radiation are based on principles of physics.
d. Predictions of the life expectancy of patients receiving radiation therapy are based on physics.

1.2 The Scientific Methods

43. The free-electron model of metals explains some of the important behaviors of metals by assuming the metal's electrons move freely through the metal without repelling one another. In what sense is the free-electron theory based on a model?
a. Its use requires constructing replicas of the metal wire in the lab.
b. It involves analyzing an imaginary system simpler than the real wire it resembles.
c. It examines a model, or ideal, behavior that other metals should imitate.
d. It attempts to examine the metal in a very realistic, or model, way.

44. A scientist wishes to study the motion of about 1,000 molecules of gas in a container by modeling them as tiny billiard balls bouncing randomly off one another. Which of the following is needed to calculate and store data on their detailed motion?
a. a group of hypotheses that cannot be practically tested in real life

b. a computer that can store and perform calculations on large data sets
c. a large amount of experimental results on the molecules and their motion
d. a collection of hypotheses that have not yet been tested regarding the molecules

45. When a large body of experimental evidence supports a hypothesis, what may the hypothesis eventually be considered?
a. observation
b. insight
c. conclusion
d. law

46. While watching some ants outside of your house, you notice that the worker ants gather in a specific area on your lawn. Which of the following is a testable hypothesis that attempts to explain why the ants gather in that specific area on the lawn.
a. The worker thought it was a nice location.
b. because ants may have to find a spot for the queen to lay eggs
c. because there may be some food particles lying there
d. because the worker ants are supposed to group together at a place.

1.3 The Language of Physics: Physical Quantities and Units

47. Which of the following would describe a length that is 2.0×10^{-3} of a meter?
a. 2.0 kilometers
b. 2.0 megameters
c. 2.0 millimeters
d. 2.0 micrometers

48. Suppose that a bathroom scale reads a person's mass as 65 kg with a 3 percent uncertainty. What is the uncertainty in their mass in kilograms?
a. a. 2 kg
b. b. 98 kg
c. c. 5 kg
d. d. 0

49. Which of the following best describes a variable?
a. a trend that shows an exponential relationship
b. something whose value can change over multiple measurements
c. a measure of how much a plot line changes along the y-axis
d. something that remains constant over multiple measurements

50. A high school track coach has just purchased a new stopwatch that has an uncertainty of ±0.05 s . Runners on the team regularly clock 100-m sprints in 12.49 s to 15.01 s . At the school's last track meet, the first-place sprinter came in at 12.04 s and the second-place sprinter came in at 12.07 s . Will the coach's new stopwatch be helpful in timing the sprint team? Why or why not?
a. No, the uncertainty in the stopwatch is too large to effectively differentiate between the sprint times.
b. No, the uncertainty in the stopwatch is too small to effectively differentiate between the sprint times.
c. Yes, the uncertainty in the stopwatch is too large to effectively differentiate between the sprint times.
d. Yes, the uncertainty in the stopwatch is too small to effectively differentiate between the sprint times.

Short Answer

1.1 Physics: Definitions and Applications

51. Describe the aims of physics.
a. Physics aims to explain the fundamental aspects of our universe and how these aspects interact with one another.
b. Physics aims to explain the biological aspects of our universe and how these aspects interact with one another.
c. Physics aims to explain the composition, structure and changes in matter occurring in the universe.
d. Physics aims to explain the social behavior of living beings in the universe.

52. Define the fields of magnetism and electricity and state how are they are related.
a. Magnetism describes the attractive force between a magnetized object and a metal like iron. Electricity involves the study of electric charges and their movements. Magnetism is not related to the electricity.
b. Magnetism describes the attractive force between a magnetized object and a metal like iron. Electricity involves the study of electric charges and their movements. Magnetism is produced by a flow electrical charges.
c. Magnetism involves the study of electric charges and their movements. Electricity describes the attractive force between a magnetized object and a metal. Magnetism is not related to the electricity.
d. Magnetism involves the study of electric charges and their movements. Electricity describes the attractive force between a magnetized object and a metal. Magnetism is produced by the flow electrical charges.

53. Describe what two topics physicists are trying to unify with relativistic quantum mechanics. How will this unification create a greater understanding of our universe?
    a. Relativistic quantum mechanics unifies quantum mechanics with Einstein's theory of relativity. The unified theory creates a greater understanding of our universe because it can explain objects of all sizes and masses.
    b. Relativistic quantum mechanics unifies classical mechanics with Einstein's theory of relativity. The unified theory creates a greater understanding of our universe because it can explain objects of all sizes and masses.
    c. Relativistic quantum mechanics unifies quantum mechanics with Einstein's theory of relativity. The unified theory creates a greater understanding of our universe because it is unable to explain objects of all sizes and masses.
    d. Relativistic quantum mechanics unifies classical mechanics with the Einstein's theory of relativity. The unified theory creates a greater understanding of our universe because it is unable to explain objects of all sizes and masses.

54. The findings of studies in quantum mechanics have been described as strange or weird compared to those of classical physics. Explain why this would be so.
    a. It is because the phenomena it explains are outside the normal range of human experience which deals with much larger objects.
    b. It is because the phenomena it explains can be perceived easily, namely, ordinary-sized objects.
    c. It is because the phenomena it explains are outside the normal range of human experience, namely, the very large and the very fast objects.
    d. It is because the phenomena it explains can be perceived easily, namely, the very large and the very fast objects.

55. How could knowledge of physics help you find a faster way to drive from your house to your school?
    a. Physics can explain the traffic on a particular street and help us know about the traffic in advance.
    b. Physics can explain about the ongoing construction of roads on a particular street and help us know about delays in the traffic in advance.
    c. Physics can explain distances, speed limits on a particular street and help us categorize faster routes.
    d. Physics can explain the closing of a particular street and help us categorize faster routes.

56. How could knowledge of physics help you build a sound and energy-efficient house?
    a. An understanding of force, pressure, heat, electricity, etc., which all involve physics, will help me design a sound and energy-efficient house.
    b. An understanding of the air composition, chemical composition of matter, etc., which all involves physics, will help me design a sound and energy-efficient house.
    c. An understanding of material cost and economic factors involving physics will help me design a sound and energy-efficient house.
    d. An understanding of geographical location and social environment which involves physics will help me design a sound and energy-efficient house.

57. What aspects of physics would a chemist likely study in trying to discover a new chemical reaction?
    a. Physics is involved in understanding whether the reactants and products dissolve in water.
    b. Physics is involved in understanding the amount of energy released or required in a chemical reaction.
    c. Physics is involved in what the products of the reaction will be.
    d. Physics is involved in understanding the types of ions produced in a chemical reaction.

1.2 The Scientific Methods

58. You notice that it takes more force to get a large box to start sliding across the floor than it takes to get the box sliding faster once it is already moving. Create a testable hypothesis that attempts to explain this observation.
    a. The floor has greater distortions of space-time for moving the sliding box faster than for the box at rest.
    b. The floor has greater distortions of space-time for the box at rest than for the sliding box.
    c. The resistance between the floor and the box is less when the box is sliding then when the box is at rest.
    d. The floor dislikes having objects move across it and therefore holds the box rigidly in place until it cannot resist the force.

59. Design an experiment that will test the following hypothesis: driving on a gravel road causes greater damage to a car than driving on a dirt road.
    a. To test the hypothesis, compare the damage to the car by driving it on a smooth road and a gravel road.
    b. To test the hypothesis, compare the damage to the car by driving it on a smooth road and a dirt road.
    c. To test the hypothesis, compare the damage to the car by driving it on a gravel road and the dirt road.
    d. This is not a testable hypothesis.

60. How is a physical model, such as a spherical mass held

in place by springs, used to represent an atom vibrating in a solid, similar to a computer-based model, such as that predicting how gravity affects the orbits of the planets?

a. Both a physical model and a computer-based model should be built around a hypothesis and could be able to test the hypothesis.
b. Both a physical model and a computer-based model should be built around a hypothesis but they cannot be used to test the hypothesis.
c. Both a physical model and a computer-based model should be built around the results of scientific studies and could be used to make predictions about the system under study.
d. Both a physical model and a computer-based model should be built around the results of scientific studies but cannot be used to make predictions about the system under study.

61. Explain the advantages and disadvantages of using a model to predict a life-or-death situation, such as whether or not an asteroid will strike Earth.

a. The advantage of using a model is that it provides predictions quickly, but the disadvantage of using a model is that it could make erroneous predictions.
b. The advantage of using a model is that it provides accurate predictions, but the disadvantage of using a model is that it takes a long time to make predictions.
c. The advantage of using a model is that it provides predictions quickly without any error. There are no disadvantages of using a scientific model.
d. The disadvantage of using models is that it takes longer time to make predictions and the predictions are inaccurate. There are no advantages to using a scientific model.

62. A friend tells you that a scientific law cannot be changed. State whether or not your friend is correct and then briefly explain your answer.

a. Correct, because laws are theories that have been proved true.
b. Correct, because theories are laws that have been proved true.
c. Incorrect, because a law is changed if new evidence contradicts it.
d. Incorrect, because a law is changed when a theory contradicts it.

63. How does a scientific law compare to a local law, such as that governing parking at your school, in terms of whether or not laws can be changed, and how universal a law is?

a. A local law applies only in a specific area, but a scientific law is applicable throughout the universe. Both the local law and the scientific law can change.
b. A local law applies only in a specific area, but a scientific law is applicable throughout the universe. A local law can change, but a scientific law cannot be changed.
c. A local law applies throughout the universe but a scientific law is applicable only in a specific area. Both the local and the scientific law can change.
d. A local law applies throughout the universe, but a scientific law is applicable only in a specific area. A local law can change, but a scientific law cannot be changed.

64. Can the validity of a model be limited, or must it be universally valid? How does this compare to the required validity of a theory or a law?

a. Models, theories and laws must be universally valid.
b. Models, theories, and laws have only limited validity.
c. Models have limited validity while theories and laws are universally valid.
d. Models and theories have limited validity while laws are universally valid.

1.3 The Language of Physics: Physical Quantities and Units

65. The speed of sound is measured at 342 m/s on a certain day. What is this in km/h? Report your answer in scientific notation.

a. 1.23×10^4 km/h
b. 1.23×10^3 km/h
c. 9.5×10^1 km/h
d. 2.05×10^{-1} km/h

66. Describe the main difference between the metric system and the U.S. Customary System.

a. In the metric system, unit changes are based on powers of 10, while in the U.S. customary system, each unit conversion has unrelated conversion factors.
b. In the metric system, each unit conversion has unrelated conversion factors, while in the U.S. customary system, unit changes are based on powers of 10.
c. In the metric system, unit changes are based on powers of 2, while in the U.S. customary system, each unit conversion has unrelated conversion factors.
d. In the metric system, each unit conversion has unrelated conversion factors, while in the U.S. customary system, unit changes are based on

powers of 2.

67. An infant's pulse rate is measured to be 130 ± 5 beats/min. What is the percent uncertainty in this measurement?

a. 2%
b. 3%
c. 4%
d. 5%

68. Explain how the uncertainty of a measurement relates to the accuracy and precision of the measuring device. Include the definitions of accuracy and precision in your answer.

a. A decrease in the precision of a measurement increases the uncertainty of the measurement, while a decrease in accuracy does not.
b. A decrease in either the precision or accuracy of a measurement increases the uncertainty of the measurement.
c. An increase in either the precision or accuracy of a measurement will increase the uncertainty of that measurement.
d. An increase in the accuracy of a measurement will increase the uncertainty of that measurement, while an increase in precision will not.

69. Describe all of the characteristics that can be determined about a straight line with a slope of -3 and a y-intercept of 50 on a graph.

a. Based on the information, the line has a negative slope. Because its y-intercept is 50 and its slope is negative, this line gradually rises on the graph as the x-value increases.
b. Based on the information, the line has a negative slope. Because its y-intercept is 50 and its slope is negative, this line gradually moves downward on the graph as the x-value increases.
c. Based on the information, the line has a positive slope. Because its y-intercept is 50 and its slope is positive, this line gradually rises on the graph as the x-value increases.
d. Based on the information, the line has a positive slope. Because its y-intercept is 50 and its slope is positive, this line gradually moves downward on the graph as the x-value increases.

70. The graph shows the temperature change over time of a heated cup of water.

What is the slope of the graph between the time period 2 min and 5 min?

a. −15 °C/min
b. −0.07 °C/min
c. 0.07 °C/min
d. 15 °C/min

Extended Response

1.2 The Scientific Methods

71. You wish to perform an experiment on the stopping distance of your new car. Create a specific experiment to measure the distance. Be sure to specifically state how you will set up and take data during your experiment.

a. Drive the car at exactly 50 mph and then press harder on the accelerator pedal until the velocity reaches the speed 60 mph and record the distance this takes.
b. Drive the car at exactly 50 mph and then apply the brakes until it stops and record the distance this takes.
c. Drive the car at exactly 50 mph and then apply the brakes until it stops and record the time it takes.
d. Drive the car at exactly 50 mph and then apply the accelerator until it reaches the speed of 60 mph and record the time it takes.

72. You wish to make a model showing how traffic flows around your city or local area. Describe the steps you would take to construct your model as well as some hypotheses that your model could test and the model's limitations in terms of what could not be tested.

a. 1. Testable hypotheses like the gravitational pull on each vehicle while in motion and the average speed of vehicles is 40 mph
   2. Non-testable hypotheses like the average number of vehicles passing is 935 per day and carbon emission from each of the moving vehicle

b. 1. Testable hypotheses like the average number of vehicles passing is 935 per day and the average speed of vehicles is 40 mph
   2. Non-testable hypotheses like the gravitational pull on each vehicle while in motion and the carbon emission from each of the moving vehicle

c. 1. Testable hypotheses like the average number of vehicles passing is 935 per day and the carbon emission from each of the moving vehicle
   2. Non-testable hypotheses like the gravitational pull on each vehicle while in motion and the average speed of the vehicles is 40 mph

d. 1. Testable hypotheses like the average number of vehicles passing is 935 per day and the gravitational pull on each vehicle while in motion
   2. Non-testable hypotheses like the average speed of vehicles is 40 mph and the carbon emission from each of the moving vehicle

73. What would play the most important role in leading to an experiment in the scientific world becoming a scientific law?

a. Further testing would need to show it is a universally followed rule.
b. The observation would have to be described in a published scientific article.
c. The experiment would have to be repeated once or twice.
d. The observer would need to be a well-known scientist whose authority was accepted.

1.3 The Language of Physics: Physical Quantities and Units

74. Tectonic plates are large segments of the Earth's crust that move slowly. Suppose that one such plate has an average speed of $4.0\ \text{cm/year}$. What distance does it move in 1.0 s at this speed? What is its speed in kilometers per million years? Report all of your answers using scientific notation.

a. 1.3×10^{-9} m; 4.0×10^{1} km/million years
b. 1.3×10^{-6} m; 4.0×10^{1} km/million years
c. 1.3×10^{-9} m; 4.0×10^{-11} km/million years
d. 1.3×10^{-6} m; 4.0×10^{-11} km/million years

75. At x = 3, a function f(x) has a positive value, with a positive slope that is decreasing in magnitude with increasing x. Which option could correspond to f(x)?

a. $y = 13x$
b. $y = x^2$
c. $y = 2x + 9$
d. $y = \frac{x}{2} + 9$

CHAPTER 2
Motion in One Dimension

Figure 2.1 Shanghai Maglev. At this rate, a train traveling from Boston to Washington, DC, a distance of 439 miles, could make the trip in under an hour and a half. Presently, the fastest train on this route takes over six hours to cover this distance. (Alex Needham, Public Domain)

Chapter Outline

2.1 Relative Motion, Distance, and Displacement

2.2 Speed and Velocity

2.3 Position vs. Time Graphs

2.4 Velocity vs. Time Graphs

INTRODUCTION Unless you have flown in an airplane, you have probably never traveled faster than 150 mph. Can you imagine traveling in a train like the one shown in Figure 2.1 that goes over 300 mph? Despite the high speed, the people riding in this train may not notice that they are moving at all unless they look out the window! This is because motion, even motion at 300 mph, is relative to the observer.

In this chapter, you will learn why it is important to identify a reference frame in order to clearly describe motion. For now, the motion you describe will be one-dimensional. Within this context, you will learn the difference between distance and displacement as well as the difference between speed and velocity. Then you will look at some graphing and problem-solving techniques.

2.1 Relative Motion, Distance, and Displacement

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Describe motion in different reference frames
- Define distance and displacement, and distinguish between the two
- Solve problems involving distance and displacement

Section Key Terms

displacement	distance	kinematics	magnitude
position	reference frame	scalar	vector

Defining Motion

Our study of physics opens with **kinematics**—the study of motion without considering its causes. Objects are in motion everywhere you look. Everything from a tennis game to a space-probe flyby of the planet Neptune involves motion. When you are resting, your heart moves blood through your veins. Even in inanimate objects, atoms are always moving.

How do you know something is moving? The location of an object at any particular time is its **position**. More precisely, you need to specify its position relative to a convenient **reference frame**. Earth is often used as a reference frame, and we often describe the position of an object as it relates to stationary objects in that reference frame. For example, a rocket launch would be described in terms of the position of the rocket with respect to Earth as a whole, while a professor's position could be described in terms of where she is in relation to the nearby white board. In other cases, we use reference frames that are not stationary but are in motion relative to Earth. To describe the position of a person in an airplane, for example, we use the airplane, not Earth, as the reference frame. (See Figure 2.2.) Thus, you can only know how fast and in what direction an object's position is changing against a background of something else that is either not moving or moving with a known speed and direction. The reference frame is the coordinate system from which the positions of objects are described.

Figure 2.2 Are clouds a useful reference frame for airplane passengers? Why or why not? (Paul Brennan, Public Domain)

Your classroom can be used as a reference frame. In the classroom, the walls are not moving. Your motion as you walk to the door, can be measured against the stationary background of the classroom walls. You can also tell if other things in the classroom are moving, such as your classmates entering the classroom or a book falling off a desk. You can also tell in what direction something is moving in the classroom. You might say, "The teacher is moving toward the door." Your reference frame allows you to determine not only that something is moving but also the direction of motion.

You could also serve as a reference frame for others' movement. If you remained seated as your classmates left the room, you would measure their movement away from your stationary location. If you and your classmates left the room together, then your perspective of their motion would be change. You, as the reference frame, would be moving in the same direction as your other moving classmates. As you will learn in the **Snap Lab**, your description of motion can be quite different when viewed from different reference frames.

Snap Lab

Looking at Motion from Two Reference Frames

In this activity you will look at motion from two reference frames. Which reference frame is correct?

- Choose an open location with lots of space to spread out so there is less chance of tripping or falling due to a collision and/or loose basketballs.
- 1 basketball

Procedure

1. Work with a partner. Stand a couple of meters away from your partner. Have your partner turn to the side so that you are looking at your partner's profile. Have your partner begin bouncing the basketball while standing in place. Describe the motion of the ball.
2. Next, have your partner again bounce the ball, but this time your partner should walk forward with the bouncing ball. You will remain stationary. Describe the ball's motion.
3. Again have your partner walk forward with the bouncing ball. This time, you should move alongside your partner while continuing to view your partner's profile. Describe the ball's motion.
4. Switch places with your partner, and repeat Steps 1–3.

GRASP CHECK

How do the different reference frames affect how you describe the motion of the ball?

a. The motion of the ball is independent of the reference frame and is same for different reference frames.
b. The motion of the ball is independent of the reference frame and is different for different reference frames.
c. The motion of the ball is dependent on the reference frame and is same for different reference frames.
d. The motion of the ball is dependent on the reference frames and is different for different reference frames.

LINKS TO PHYSICS

History: Galileo's Ship

Figure 2.3 Galileo Galilei (1564–1642) studied motion and developed the concept of a reference frame. (Domenico Tintoretto)

The idea that a description of motion depends on the reference frame of the observer has been known for hundreds of years. The 17th-century astronomer Galileo Galilei (Figure 2.3) was one of the first scientists to explore this idea. Galileo suggested the following thought experiment: Imagine a windowless ship moving at a constant speed and direction along a perfectly calm sea. Is there a way that a person inside the ship can determine whether the ship is moving? You can extend this thought experiment

by also imagining a person standing on the shore. How can a person on the shore determine whether the ship is moving?

Galileo came to an amazing conclusion. Only by looking at each other can a person in the ship or a person on shore describe the motion of one relative to the other. In addition, their descriptions of motion would be identical. A person inside the ship would describe the person on the land as moving past the ship. The person on shore would describe the ship and the person inside it as moving past. Galileo realized that observers moving at a constant speed and direction relative to each other describe motion in the same way. Galileo had discovered that a description of motion is only meaningful if you specify a reference frame.

GRASP CHECK

Imagine standing on a platform watching a train pass by. According to Galileo's conclusions, how would your description of motion and the description of motion by a person riding on the train compare?

a. I would see the train as moving past me, and a person on the train would see me as stationary.
b. I would see the train as moving past me, and a person on the train would see me as moving past the train.
c. I would see the train as stationary, and a person on the train would see me as moving past the train.
d. I would see the train as stationary, and a person on the train would also see me as stationary.

Distance vs. Displacement

As we study the motion of objects, we must first be able to describe the object's position. Before your parent drives you to school, the car is sitting in your driveway. Your driveway is the starting position for the car. When you reach your high school, the car has changed position. Its new position is your school.

Figure 2.4 Your total change in position is measured from your house to your school.

Physicists use variables to represent terms. We will use **d** to represent car's position. We will use a subscript to differentiate between the initial position, $\mathbf{d}_o$, and the final position, $\mathbf{d}_f$. In addition, vectors, which we will discuss later, will be in bold or will have an arrow above the variable. Scalars will be italicized.

TIPS FOR SUCCESS

In some books, **x** or **s** is used instead of **d** to describe position. In $\mathbf{d}_o$, said *d naught*, the subscript o stands for *initial*. When we begin to talk about two-dimensional motion, sometimes other subscripts will be used to describe horizontal position, $\mathbf{d}_x$, or vertical position, $\mathbf{d}_y$. So, you might see references to $\mathbf{d}_{ox}$ and $\mathbf{d}_{fy}$.

Now imagine driving from your house to a friend's house located several kilometers away. How far would you drive? The **distance** an object moves is the length of the path between its initial position and its final position. The distance you drive to your friend's house depends on your path. As shown in Figure 2.5, distance is different from the length of a straight line between two points. The distance you drive to your friend's house is probably longer than the straight line between the two houses.

Figure 2.5 A short line separates the starting and ending points of this motion, but the distance along the path of motion is considerably longer.

We often want to be more precise when we talk about position. The description of an object's motion often includes more than just the distance it moves. For instance, if it is a five kilometer drive to school, the distance traveled is 5 kilometers. After dropping you off at school and driving back home, your parent will have traveled a total distance of 10 kilometers. The car and your parent will end up in the same starting position in space. The net change in position of an object is its **displacement**, or $\Delta\mathbf{d}$. The Greek letter delta, Δ, means *change in*.

Figure 2.6 The total distance that your car travels is 10 km, but the total displacement is 0.

Snap Lab

Distance vs. Displacement

In this activity you will compare distance and displacement. Which term is more useful when making measurements?

- 1 recorded song available on a portable device
- 1 tape measure
- 3 pieces of masking tape
- A room (like a gym) with a wall that is large and clear enough for all pairs of students to walk back and forth without running into each other.

Procedure

1. One student from each pair should stand with their back to the longest wall in the classroom. Students should stand at least 0.5 meters away from each other. Mark this starting point with a piece of masking tape.
2. The second student from each pair should stand facing their partner, about two to three meters away. Mark this point

with a second piece of masking tape.
3. Student pairs line up at the starting point along the wall.
4. The teacher turns on the music. Each pair walks back and forth from the wall to the second marked point until the music stops playing. Keep count of the number of times you walk across the floor.
5. When the music stops, mark your ending position with the third piece of masking tape.
6. Measure from your starting, initial position to your ending, final position.
7. Measure the length of your path from the starting position to the second marked position. Multiply this measurement by the total number of times you walked across the floor. Then add this number to your measurement from step 6.
8. Compare the two measurements from steps 6 and 7.

GRASP CHECK

1. Which measurement is your total distance traveled?
2. Which measurement is your displacement?
3. When might you want to use one over the other?

a. Measurement of the total length of your path from the starting position to the final position gives the distance traveled, and the measurement from your initial position to your final position is the displacement. Use distance to describe the total path between starting and ending points,and use displacement to describe the shortest path between starting and ending points.
b. Measurement of the total length of your path from the starting position to the final position is distance traveled, and the measurement from your initial position to your final position is displacement. Use distance to describe the shortest path between starting and ending points, and use displacement to describe the total path between starting and ending points.
c. Measurement from your initial position to your final position is distance traveled, and the measurement of the total length of your path from the starting position to the final position is displacement. Use distance to describe the total path between starting and ending points, and use displacement to describe the shortest path between starting and ending points.
d. Measurement from your initial position to your final position is distance traveled, and the measurement of the total length of your path from the starting position to the final position is displacement. Use distance to describe the shortest path between starting and ending points, and use displacement to describe the total path between starting and ending points.

If you are describing only your drive to school, then the distance traveled and the displacement are the same—5 kilometers. When you are describing the entire round trip, distance and displacement are different. When you describe distance, you only include the **magnitude**, the size or amount, of the distance traveled. However, when you describe the displacement, you take into account both the magnitude of the change in position and the direction of movement.

In our previous example, the car travels a total of 10 kilometers, but it drives five of those kilometers forward toward school and five of those kilometers back in the opposite direction. If we ascribe the forward direction a positive (+) and the opposite direction a negative (–), then the two quantities will cancel each other out when added together.

A quantity, such as distance, that has magnitude (i.e., how big or how much) but does not take into account direction is called a **scalar**. A quantity, such as displacement, that has both magnitude and direction is called a **vector**.

WATCH PHYSICS

Vectors & Scalars

This video (http://openstax.org/l/28vectorscalar) introduces and differentiates between vectors and scalars. It also introduces quantities that we will be working with during the study of kinematics.

Click to view content (https://www.khanacademy.org/embed_video?v=ihNZlp7iUHE)

GRASP CHECK

How does this video (https://www.khanacademy.org/science/ap-physics-1/ap-one-dimensional-motion/ap-physics-foundations/v/introduction-to-vectors-and-scalars) help you understand the difference between distance and displacement? Describe the differences between vectors and scalars using physical quantities as examples.

a. It explains that distance is a vector and direction is important, whereas displacement is a scalar and it has no direction attached to it.
b. It explains that distance is a scalar and direction is important, whereas displacement is a vector and it has no direction attached to it.
c. It explains that distance is a scalar and it has no direction attached to it, whereas displacement is a vector and direction is important.
d. It explains that both distance and displacement are scalar and no directions are attached to them.

Displacement Problems

Hopefully you now understand the conceptual difference between distance and displacement. Understanding concepts is half the battle in physics. The other half is math. A stumbling block to new physics students is trying to wade through the math of physics while also trying to understand the associated concepts. This struggle may lead to misconceptions and answers that make no sense. Once the concept is mastered, the math is far less confusing.

So let's review and see if we can make sense of displacement in terms of numbers and equations. You can calculate an object's displacement by subtracting its original position, $\mathbf{d_o}$, from its final position $\mathbf{d_f}$. In math terms that means

$$\Delta\mathbf{d} = \mathbf{d}_f - \mathbf{d}_0.$$

If the final position is the same as the initial position, then $\Delta\mathbf{d} = 0$.

To assign numbers and/or direction to these quantities, we need to define an axis with a positive and a negative direction. We also need to define an origin, or *O*. In Figure 2.6, the axis is in a straight line with home at zero and school in the positive direction. If we left home and drove the opposite way from school, motion would have been in the negative direction. We would have assigned it a negative value. In the round-trip drive, $\mathbf{d}_f$ and $\mathbf{d}_o$ were both at zero kilometers. In the one way trip to school, $\mathbf{d}_f$ was at 5 kilometers and $\mathbf{d}_o$ was at zero km. So, $\Delta\mathbf{d}$ was 5 kilometers.

TIPS FOR SUCCESS

You may place your origin wherever you would like. You have to make sure that you calculate all distances consistently from your zero and you define one direction as positive and the other as negative. Therefore, it makes sense to choose the easiest axis, direction, and zero. In the example above, we took home to be zero because it allowed us to avoid having to interpret a solution with a negative sign.

WORKED EXAMPLE

Calculating Distance and Displacement

A cyclist rides 3 km west and then turns around and rides 2 km east. (a) What is her displacement? (b) What distance does she ride? (c) What is the magnitude of her displacement?

Strategy

To solve this problem, we need to find the difference between the final position and the initial position while taking care to note the direction on the axis. The final position is the sum of the two displacements, $\Delta\mathbf{d}_1$ and $\Delta\mathbf{d}_2$.

Solution

a. Displacement: The rider's displacement is $\Delta\mathbf{d} = \mathbf{d}_f - \mathbf{d}_0 = -1 \text{ km}$.
b. Distance: The distance traveled is 3 km + 2 km = 5 km.
c. The magnitude of the displacement is 1 km.

Discussion

The displacement is negative because we chose east to be positive and west to be negative. We could also have described the displacement as 1 km west. When calculating displacement, the direction mattered, but when calculating distance, the direction did not matter. The problem would work the same way if the problem were in the north–south or *y*-direction.

TIPS FOR SUCCESS

Physicists like to use standard units so it is easier to compare notes. The standard units for calculations are called *SI* units (International System of Units). SI units are based on the metric system. The SI unit for displacement is the meter (m), but sometimes you will see a problem with kilometers, miles, feet, or other units of length. If one unit in a problem is an SI unit and another is not, you will need to convert all of your quantities to the same system before you can carry out the calculation.

Practice Problems

1. On an axis in which moving from right to left is positive, what is the displacement and distance of a student who walks 32 m to the right and then 17 m to the left?
   a. Displacement is -15 m and distance is -49 *m*.
   b. Displacement is -15 m and distance is 49 m.
   c. Displacement is 15 m and distance is -49 m.
   d. Displacement is 15 m and distance is 49 m.

2. Tiana jogs 1.5 km along a straight path and then turns and jogs 2.4 km in the opposite direction. She then turns back and jogs 0.7 km in the original direction. Let Tiana's original direction be the positive direction. What are the displacement and distance she jogged?
   a. Displacement is 4.6 km,and distance is -0.2 km.
   b. Displacement is -0.2 km, and distance is 4.6 km.
   c. Displacement is 4.6 km, and distance is +0.2 km.
   d. Displacement is +0.2 km, and distance is 4.6 km.

WORK IN PHYSICS

Mars Probe Explosion

Figure 2.7 The Mars Climate Orbiter disaster illustrates the importance of using the correct calculations in physics. (NASA)

Physicists make calculations all the time, but they do not always get the right answers. In 1998, NASA, the National Aeronautics and Space Administration, launched the Mars Climate Orbiter, shown in Figure 2.7, a $125-million-dollar satellite designed to monitor the Martian atmosphere. It was supposed to orbit the planet and take readings from a safe distance. The American scientists made calculations in English units (feet, inches, pounds, etc.) and forgot to convert their answers to the standard metric SI units. This was a very costly mistake. Instead of orbiting the planet as planned, the Mars Climate Orbiter ended up flying into the Martian atmosphere. The probe disintegrated. It was one of the biggest embarrassments in NASA's history.

GRASP CHECK

In 1999 the Mars Climate Orbiter crashed because calculation were performed in English units instead of SI units. At one point the orbiter was just 187,000 feet above the surface, which was too close to stay in orbit. What was the height of the orbiter at this time in kilometers? (Assume 1 meter equals 3.281 feet.)

a. 16 km
b. 18 km
c. 57 km
d. 614 km

Check Your Understanding

3. What does it mean when motion is described as relative?
   a. It means that motion of any object is described relative to the motion of Earth.
   b. It means that motion of any object is described relative to the motion of any other object.
   c. It means that motion is independent of the frame of reference.
   d. It means that motion depends on the frame of reference selected.

4. If you and a friend are standing side-by-side watching a soccer game, would you both view the motion from the same reference frame?
   a. Yes, we would both view the motion from the same reference point because both of us are at rest in Earth's frame of reference.
   b. Yes, we would both view the motion from the same reference point because both of us are observing the motion from two points on the same straight line.
   c. No, we would both view the motion from different reference points because motion is viewed from two different points; the reference frames are similar but not the same.
   d. No, we would both view the motion from different reference points because response times may be different; so, the motion observed by both of us would be different.

5. What is the difference between distance and displacement?
   a. Distance has both magnitude and direction, while displacement has magnitude but no direction.
   b. Distance has magnitude but no direction, while displacement has both magnitude and direction.
   c. Distance has magnitude but no direction, while displacement has only direction.
   d. There is no difference. Both distance and displacement have magnitude and direction.

6. Which situation correctly identifies a race car's distance traveled and the magnitude of displacement during a one-lap car race?
   a. The perimeter of the race track is the distance, and the shortest distance between the start line and the finish line is the magnitude of displacement.
   b. The perimeter of the race track is the magnitude of displacement, and the shortest distance between the start and finish line is the distance.
   c. The perimeter of the race track is both the distance and magnitude of displacement.
   d. The shortest distance between the start line and the finish line is both the distance and magnitude of displacement.

7. Why is it important to specify a reference frame when describing motion?
   a. Because Earth is continuously in motion; an object at rest on Earth will be in motion when viewed from outer space.
   b. Because the position of a moving object can be defined only when there is a fixed reference frame.

c. Because motion is a relative term; it appears differently when viewed from different reference frames.
d. Because motion is always described in Earth's frame of reference; if another frame is used, it has to be specified with each situation.

2.2 Speed and Velocity

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Calculate the average speed of an object
- Relate displacement and average velocity

Section Key Terms

average speed	average velocity	instantaneous speed
instantaneous velocity	speed	velocity

Speed

There is more to motion than distance and displacement. Questions such as, "How long does a foot race take?" and "What was the runner's speed?" cannot be answered without an understanding of other concepts. In this section we will look at time, speed, and velocity to expand our understanding of motion.

A description of how fast or slow an object moves is its speed. **Speed** is the rate at which an object changes its location. Like distance, speed is a scalar because it has a magnitude but not a direction. Because speed is a rate, it depends on the time interval of motion. You can calculate the elapsed time or the change in time, Δt, of motion as the difference between the ending time and the beginning time

$$\Delta t = t_f - t_0.$$

The SI unit of time is the second (s), and the SI unit of speed is meters per second (m/s), but sometimes kilometers per hour (km/h), miles per hour (mph) or other units of speed are used.

When you describe an object's speed, you often describe the average over a time period. **Average speed**, v_{avg}, is the distance traveled divided by the time during which the motion occurs.

$$v_{avg} = \frac{\text{distance}}{\text{time}}$$

You can, of course, rearrange the equation to solve for either distance or time

$$\text{time} = \frac{\text{distance}}{v_{avg}}.$$

$$\text{distance} = v_{avg} \times \text{time}$$

Suppose, for example, a car travels 150 kilometers in 3.2 hours. Its average speed for the trip is

$$\begin{aligned} v_{avg} &= \frac{\text{distance}}{\text{time}} \\ &= \frac{150\ \text{km}}{3.2\ \text{h}} \\ &= 47\ \text{km/h}. \end{aligned}$$

A car's speed would likely increase and decrease many times over a 3.2 hour trip. Its speed at a specific instant in time, however, is its **instantaneous speed**. A car's speedometer describes its instantaneous speed.

Figure 2.8 During a 30-minute round trip to the store, the total distance traveled is 6 km. The average speed is 12 km/h. The displacement for the round trip is zero, because there was no net change in position.

WORKED EXAMPLE

Calculating Average Speed

A marble rolls 5.2 m in 1.8 s. What was the marble's average speed?

Strategy

We know the distance the marble travels, 5.2 m, and the time interval, 1.8 s. We can use these values in the average speed equation.

Solution

$$v_{avg} = \frac{\text{distance}}{\text{time}} = \frac{5.2\ \text{m}}{1.8\ \text{s}} = 2.9\ \text{m/s}$$

Discussion

Average speed is a scalar, so we do not include direction in the answer. We can check the reasonableness of the answer by estimating: 5 meters divided by 2 seconds is 2.5 m/s. Since 2.5 m/s is close to 2.9 m/s, the answer is reasonable. This is about the speed of a brisk walk, so it also makes sense.

Practice Problems

8. A pitcher throws a baseball from the pitcher's mound to home plate in 0.46 s. The distance is 18.4 m. What was the average speed of the baseball?
   a. 40 m/s
   b. - 40 m/s
   c. 0.03 m/s
   d. 8.5 m/s

9. Cassie walked to her friend's house with an average speed of 1.40 m/s. The distance between the houses is 205 m. How long did the trip take her?
   a. 146 s
   b. 0.01 s
   c. 2.50 min
   d. 287 s

Velocity

The vector version of speed is velocity. **Velocity** describes the speed and direction of an object. As with speed, it is useful to describe either the average velocity over a time period or the velocity at a specific moment. **Average velocity** is displacement divided by the time over which the displacement occurs.

$$v_{avg} = \frac{\text{distance}}{\text{time}} = \frac{\Delta \mathbf{d}}{\Delta t} = \frac{\mathbf{d}_f - \mathbf{d}_0}{t_f - t_0}$$

Velocity, like speed, has SI units of meters per second (m/s), but because it is a vector, you must also include a direction. Furthermore, the variable **v** for velocity is bold because it is a vector, which is in contrast to the variable *v* for speed which is italicized because it is a scalar quantity.

TIPS FOR SUCCESS

It is important to keep in mind that the average speed is not the same thing as the average velocity without its direction. Like we saw with displacement and distance in the last section, changes in direction over a time interval have a bigger effect on speed and velocity.

Suppose a passenger moved toward the back of a plane with an average velocity of −4 m/s. We cannot tell from the average velocity whether the passenger stopped momentarily or backed up before he got to the back of the plane. To get more details, we must consider smaller segments of the trip over smaller time intervals such as those shown in Figure 2.9. If you consider infinitesimally small intervals, you can define **instantaneous velocity**, which is the velocity at a specific instant in time. Instantaneous velocity and average velocity are the same if the velocity is constant.

Figure 2.9 The diagram shows a more detailed record of an airplane passenger heading toward the back of the plane, showing smaller segments of his trip.

Earlier, you have read that distance traveled can be different than the magnitude of displacement. In the same way, speed can be different than the magnitude of velocity. For example, you drive to a store and return home in half an hour. If your car's odometer shows the total distance traveled was 6 km, then your average speed was 12 km/h. Your average velocity, however, was zero because your displacement for the round trip is zero.

WATCH PHYSICS

Calculating Average Velocity or Speed

This video (http://openstax.org/l/28avgvelocity) reviews vectors and scalars and describes how to calculate average velocity and average speed when you know displacement and change in time. The video also reviews how to convert km/h to m/s.

Click to view content (https://www.khanacademy.org/embed_video?v=MAS6mBRZZXA)

GRASP CHECK

Which of the following fully describes a vector and a scalar quantity and correctly provides an example of each?

a. A scalar quantity is fully described by its magnitude, while a vector needs both magnitude and direction to fully describe it. Displacement is an example of a scalar quantity and time is an example of a vector quantity.
b. A scalar quantity is fully described by its magnitude, while a vector needs both magnitude and direction to fully describe it. Time is an example of a scalar quantity and displacement is an example of a vector quantity.
c. A scalar quantity is fully described by its magnitude and direction, while a vector needs only magnitude to fully describe it. Displacement is an example of a scalar quantity and time is an example of a vector quantity.
d. A scalar quantity is fully described by its magnitude and direction, while a vector needs only magnitude to fully describe it. Time is an example of a scalar quantity and displacement is an example of a vector quantity.

WORKED EXAMPLE

Calculating Average Velocity

A student has a displacement of 304 m north in 180 s. What was the student's average velocity?

Strategy

We know that the displacement is 304 m north and the time is 180 s. We can use the formula for average velocity to solve the problem.

Solution

$$\mathbf{v}_{\text{avg}} = \frac{\Delta \mathbf{d}}{\Delta t} = \frac{304 \text{ m}}{180 \text{ s}} = 1.7 \text{ m/s north} \tag{2.1}$$

Discussion

Since average velocity is a vector quantity, you must include direction as well as magnitude in the answer. Notice, however, that the direction can be omitted until the end to avoid cluttering the problem. Pay attention to the significant figures in the problem. The distance 304 m has three significant figures, but the time interval 180 s has only two, so the quotient should have only two significant figures.

TIPS FOR SUCCESS

Note the way scalars and vectors are represented. In this book d represents distance and displacement. Similarly, v represents speed, and v represents velocity. A variable that is not bold indicates a scalar quantity, and a bold variable indicates a vector quantity. Vectors are sometimes represented by small arrows above the variable.

WORKED EXAMPLE

Solving for Displacement when Average Velocity and Time are Known

Layla jogs with an average velocity of 2.4 m/s east. What is her displacement after 46 seconds?

Strategy

We know that Layla's average velocity is 2.4 m/s east, and the time interval is 46 seconds. We can rearrange the average velocity formula to solve for the displacement.

Solution

$$\begin{aligned} \mathbf{v}_{\text{avg}} &= \frac{\Delta \mathbf{d}}{\Delta t} \\ \Delta \mathbf{d} &= \mathbf{v}_{avg} \Delta t \\ &= (2.4 \text{ m/s})(46 \text{ s}) \\ &= 1.1 \times 10^2 \text{ m east} \end{aligned} \tag{2.2}$$

Discussion

The answer is about 110 m east, which is a reasonable displacement for slightly less than a minute of jogging. A calculator shows the answer as 110.4 m. We chose to write the answer using scientific notation because we wanted to make it clear that we only

used two significant figures.

TIPS FOR SUCCESS

Dimensional analysis is a good way to determine whether you solved a problem correctly. Write the calculation using only units to be sure they match on opposite sides of the equal mark. In the worked example, you have m = (m/s)(s). Since seconds is in the denominator for the average velocity and in the numerator for the time, the unit cancels out leaving only m and, of course, m = m.

WORKED EXAMPLE

Solving for Time when Displacement and Average Velocity are Known

Phillip walks along a straight path from his house to his school. How long will it take him to get to school if he walks 428 m west with an average velocity of 1.7 m/s west?

Strategy

We know that Phillip's displacement is 428 m west, and his average velocity is 1.7 m/s west. We can calculate the time required for the trip by rearranging the average velocity equation.

Solution

$$\begin{aligned} \mathbf{v}_{\text{avg}} &= \frac{\Delta \mathbf{d}}{\Delta t} \\ \Delta t &= \frac{\Delta \mathbf{d}}{\mathbf{v}_{\text{avg}}} \\ &= \frac{428 \text{ m}}{1.7 \text{ m/s}} \\ &= 2.5 \times 10^2 \text{ s} \end{aligned} \tag{2.3}$$

Discussion

Here again we had to use scientific notation because the answer could only have two significant figures. Since time is a scalar, the answer includes only a magnitude and not a direction.

Practice Problems

10. A trucker drives along a straight highway for 0.25 h with a displacement of 16 km south. What is the trucker's average velocity?
    a. 4 km/h north
    b. 4 km/h south
    c. 64 km/h north
    d. 64 km/h south

11. A bird flies with an average velocity of 7.5 m/s east from one branch to another in 2.4 s. It then pauses before flying with an average velocity of 6.8 m/s east for 3.5 s to another branch. What is the bird's total displacement from its starting point?
    a. 42 m west
    b. 6 m west
    c. 6 m east
    d. 42 m east

Virtual Physics

The Walking Man

In this simulation you will put your cursor on the man and move him first in one direction and then in the opposite direction. Keep the *Introduction* tab active. You can use the *Charts* tab after you learn about graphing motion later in this chapter. Carefully watch the sign of the numbers in the position and velocity boxes. Ignore the acceleration box for now. See if you can make the man's position positive while the velocity is negative. Then see if you can do the opposite.

Click to view content (https://archive.cnx.org/specials/e2ca52af-8c6b-450e-ac2f-9300b38e8739/moving-man/)

GRASP CHECK

Which situation correctly describes when the moving man's position was negative but his velocity was positive?

a. Man moving toward o from left of o
b. Man moving toward o from right of o
c. Man moving away from o from left of o
d. Man moving away from o from right of o

Check Your Understanding

12. Two runners travel along the same straight path. They start at the same time, and they end at the same time, but at the halfway mark, they have different instantaneous velocities. Is it possible for them to have the same average velocity for the trip?

a. Yes, because average velocity depends on the net or total displacement.
b. Yes, because average velocity depends on the total distance traveled.
c. No, because the velocities of both runners must remain the exactly same throughout the journey.
d. No, because the instantaneous velocities of the runners must remain same midway but can be different elsewhere.

13. If you divide the total distance traveled on a car trip (as determined by the odometer) by the time for the trip, are you calculating the average speed or the magnitude of the average velocity, and under what circumstances are these two quantities the same?

a. Average speed. Both are the same when the car is traveling at a constant speed and changing direction.
b. Average speed. Both are the same when the speed is constant and the car does not change its direction.
c. Magnitude of average velocity. Both are same when the car is traveling at a constant speed.
d. Magnitude of average velocity. Both are same when the car does not change its direction.

14. Is it possible for average velocity to be negative?

a. Yes, in cases when the net displacement is negative.
b. Yes, if the body keeps changing its direction during motion.
c. No, average velocity describes only magnitude and not the direction of motion.
d. No, average velocity describes only the magnitude in the positive direction of motion.

2.3 Position vs. Time Graphs

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Explain the meaning of slope in position vs. time graphs
- Solve problems using position vs. time graphs

Section Key Terms

dependent variable	independent variable	tangent

Graphing Position as a Function of Time

A graph, like a picture, is worth a thousand words. Graphs not only contain numerical information, they also reveal relationships between physical quantities. In this section, we will investigate kinematics by analyzing graphs of position over time.

Graphs in this text have perpendicular axes, one horizontal and the other vertical. When two physical quantities are plotted against each other, the horizontal axis is usually considered the **independent variable**, and the vertical axis is the **dependent variable**. In algebra, you would have referred to the horizontal axis as the *x*-axis and the vertical axis as the *y*-axis. As in Figure 2.10, a straight-line graph has the general form $y = mx + b$.

Here m is the slope, defined as the rise divided by the run (as seen in the figure) of the straight line. The letter b is the y-intercept which is the point at which the line crosses the vertical, y-axis. In terms of a physical situation in the real world, these quantities will take on a specific significance, as we will see below. (Figure 2.10.)

Figure 2.10 The diagram shows a straight-line graph. The equation for the straight line is y equals $mx + b$.

In physics, time is usually the independent variable. Other quantities, such as displacement, are said to depend upon it. A graph of position versus time, therefore, would have position on the vertical axis (dependent variable) and time on the horizontal axis (independent variable). In this case, to what would the slope and y-intercept refer? Let's look back at our original example when studying distance and displacement.

The drive to school was 5 km from home. Let's assume it took 10 minutes to make the drive and that your parent was driving at a constant velocity the whole time. The position versus time graph for this section of the trip would look like that shown in Figure 2.11.

Figure 2.11 A graph of position versus time for the drive to school is shown. What would the graph look like if we added the return trip?

As we said before, $\mathbf{d}_o = 0$ because we call home our O and start calculating from there. In Figure 2.11, the line starts at $\mathbf{d} = 0$, as well. This is the b in our equation for a straight line. Our initial position in a position versus time graph is always the place where the graph crosses the x-axis at $t = 0$. What is the slope? The *rise* is the change in position, (i.e., displacement) and the *run* is the change in time. This relationship can also be written

$$\frac{\Delta \mathbf{d}}{\Delta t}. \tag{2.4}$$

This relationship was how we defined average velocity. Therefore, the slope in a $\mathbf{d}$ versus t graph, is the average velocity.

TIPS FOR SUCCESS

Sometimes, as is the case where we graph both the trip to school and the return trip, the behavior of the graph looks different during different time intervals. If the graph looks like a series of straight lines, then you can calculate the average velocity for each time interval by looking at the slope. If you then want to calculate the average velocity for the entire trip, you can do a

weighted average.

Let's look at another example. Figure 2.12 shows a graph of position versus time for a jet-powered car on a very flat dry lake bed in Nevada.

Figure 2.12 The diagram shows a graph of position versus time for a jet-powered car on the Bonneville Salt Flats.

Using the relationship between dependent and independent variables, we see that the slope in the graph in Figure 2.12 is average velocity, $\mathbf{v}_{avg}$ and the intercept is displacement at time zero—that is, $\mathbf{d}_o$. Substituting these symbols into $y = mx + b$ gives

$$\mathbf{d} = \mathbf{v}t + \mathbf{d}_0 \tag{2.5}$$

or

$$\mathbf{d} = \mathbf{d}_0 + \mathbf{v}t. \tag{2.6}$$

Thus a graph of position versus time gives a general relationship among displacement, velocity, and time, as well as giving detailed numerical information about a specific situation. From the figure we can see that the car has a position of 400 m at $t = 0$ s, 650 m at $t = 1.0$ s, and so on. And we can learn about the object's velocity, as well.

Snap Lab

Graphing Motion

In this activity, you will release a ball down a ramp and graph the ball's displacement vs. time.

- Choose an open location with lots of space to spread out so there is less chance for tripping or falling due to rolling balls.

- 1 ball
- 1 board
- 2 or 3 books
- 1 stopwatch
- 1 tape measure
- 6 pieces of masking tape
- 1 piece of graph paper
- 1 pencil

Procedure

1. Build a ramp by placing one end of the board on top of the stack of books. Adjust location, as necessary, until there is no obstacle along the straight line path from the bottom of the ramp until at least the next 3 m.
2. Mark distances of 0.5 m, 1.0 m, 1.5 m, 2.0 m, 2.5 m, and 3.0 m from the bottom of the ramp. Write the distances on the tape.

3. Have one person take the role of the experimenter. This person will release the ball from the top of the ramp. If the ball does not reach the 3.0 m mark, then increase the incline of the ramp by adding another book. Repeat this Step as necessary.
4. Have the experimenter release the ball. Have a second person, the timer, begin timing the trial once the ball reaches the bottom of the ramp and stop the timing once the ball reaches 0.5 m. Have a third person, the recorder, record the time in a data table.
5. Repeat Step 4, stopping the times at the distances of 1.0 m, 1.5 m, 2.0 m, 2.5 m, and 3.0 m from the bottom of the ramp.
6. Use your measurements of time and the displacement to make a position vs. time graph of the ball's motion.
7. Repeat Steps 4 through 6, with different people taking on the roles of experimenter, timer, and recorder. Do you get the same measurement values regardless of who releases the ball, measures the time, or records the result? Discuss possible causes of discrepancies, if any.

GRASP CHECK

True or False: The average speed of the ball will be less than the average velocity of the ball.

a. True
b. False

Solving Problems Using Position vs. Time Graphs

So how do we use graphs to solve for things we want to know like velocity?

WORKED EXAMPLE

Using Position–Time Graph to Calculate Average Velocity: Jet Car

Find the average velocity of the car whose position is graphed in Figure 1.13.

Strategy

The slope of a graph of d vs. t is average velocity, since slope equals rise over run.

$$\text{slope} = \frac{\Delta \mathbf{d}}{\Delta t} = \mathbf{v} \tag{2.7}$$

Since the slope is constant here, any two points on the graph can be used to find the slope.

Solution

1. Choose two points on the line. In this case, we choose the points labeled on the graph: (6.4 s, 2000 m) and (0.50 s, 525 m). (Note, however, that you could choose any two points.)
2. Substitute the **d** and t values of the chosen points into the equation. Remember in calculating change (Δ) we always use final value minus initial value.

$$\begin{aligned} \mathbf{v} &= \frac{\Delta \mathbf{d}}{\Delta t} \\ &= \frac{2000\ \text{m} - 525\ \text{m}}{6.4\ \text{s} - 0.50\ \text{s}}, \\ &= 250\ \text{m/s} \end{aligned} \tag{2.8}$$

Discussion

This is an impressively high land speed (900 km/h, or about 560 mi/h): much greater than the typical highway speed limit of 27 m/s or 96 km/h, but considerably shy of the record of 343 m/s or 1,234 km/h, set in 1997.

But what if the graph of the position is more complicated than a straight line? What if the object speeds up or turns around and goes backward? Can we figure out anything about its velocity from a graph of that kind of motion? Let's take another look at the jet-powered car. The graph in Figure 2.13 shows its motion as it is getting up to speed after starting at rest. Time starts at zero for this motion (as if measured with a stopwatch), and the displacement and velocity are initially 200 m and 15 m/s, respectively.

Figure 2.13 The diagram shows a graph of the position of a jet-powered car during the time span when it is speeding up. The slope of a distance versus time graph is velocity. This is shown at two points. Instantaneous velocity at any point is the slope of the tangent at that point.

Figure 2.14 A U.S. Air Force jet car speeds down a track. (Matt Trostle, Flickr)

The graph of position versus time in Figure 2.13 is a curve rather than a straight line. The slope of the curve becomes steeper as time progresses, showing that the velocity is increasing over time. The slope at any point on a position-versus-time graph is the instantaneous velocity at that point. It is found by drawing a straight line **tangent** to the curve at the point of interest and taking the slope of this straight line. Tangent lines are shown for two points in Figure 2.13. The average velocity is the net displacement divided by the time traveled.

WORKED EXAMPLE

Using Position–Time Graph to Calculate Average Velocity: Jet Car, Take Two

Calculate the instantaneous velocity of the jet car at a time of 25 s by finding the slope of the tangent line at point Q in Figure 2.13.

Strategy

The slope of a curve at a point is equal to the slope of a straight line tangent to the curve at that point.

Solution

1. Find the tangent line to the curve at $t = 25 \text{ s}$.
2. Determine the endpoints of the tangent. These correspond to a position of 1,300 m at time 19 s and a position of 3120 m at time 32 s.

3. Plug these endpoints into the equation to solve for the slope, **v**.

$$\begin{aligned} \text{slope} &= v_Q = \frac{\Delta d_Q}{\Delta t_Q} \\ &= \frac{(3120-1300)\ \text{m}}{(32-19)\ \text{s}} \\ &= \frac{1820\ \text{m}}{13\ \text{s}} \\ &= 140\ \text{m/s} \end{aligned} \tag{2.9}$$

Discussion

The entire graph of **v** versus *t* can be obtained in this fashion.

Practice Problems

15. Calculate the average velocity of the object shown in the graph below over the whole time interval.

a. 0.25 m/s
b. 0.31 m/s
c. 3.2 m/s
d. 4.00 m/s

16. True or False: By taking the slope of the curve in the graph you can verify that the velocity of the jet car is 115 m/s at $t = 20$ s.

a. True
b. False

Check Your Understanding

17. Which of the following information about motion can be determined by looking at a position vs. time graph that is a straight line?

a. frame of reference
b. average acceleration
c. velocity
d. direction of force applied

18. True or False: The position vs time graph of an object that is speeding up is a straight line.
a. True
b. False

2.4 Velocity vs. Time Graphs

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Explain the meaning of slope and area in velocity vs. time graphs
- Solve problems using velocity vs. time graphs

Section Key Terms

acceleration

Graphing Velocity as a Function of Time

Earlier, we examined graphs of position versus time. Now, we are going to build on that information as we look at graphs of velocity vs. time. Velocity is the rate of change of displacement. **Acceleration** is the rate of change of velocity; we will discuss acceleration more in another chapter. These concepts are all very interrelated.

Virtual Physics

Maze Game

In this simulation you will use a vector diagram to manipulate a ball into a certain location without hitting a wall. You can manipulate the ball directly with position or by changing its velocity. Explore how these factors change the motion. If you would like, you can put it on the *a* setting, as well. This is acceleration, which measures the rate of change of velocity. We will explore acceleration in more detail later, but it might be interesting to take a look at it here.

Click to view content (https://archive.cnx.org/specials/30e37034-2fbd-11e5-83a2-03be60006ece/maze-game/)

GRASP CHECK

Click to view content (https://archive.cnx.org/specials/30e37034-2fbd-11e5-83a2-03be60006ece/maze-game/#sim-maze-game)

a. The ball can be easily manipulated with displacement because the arena is a position space.
b. The ball can be easily manipulated with velocity because the arena is a position space.
c. The ball can be easily manipulated with displacement because the arena is a velocity space.
d. The ball can be easily manipulated with velocity because the arena is a velocity space.

What can we learn about motion by looking at velocity vs. time graphs? Let's return to our drive to school, and look at a graph of position versus time as shown in Figure 2.15.

Figure 2.15 A graph of position versus time for the drive to and from school is shown.

We assumed for our original calculation that your parent drove with a constant velocity to and from school. We now know that the car could not have gone from rest to a constant velocity without speeding up. So the actual graph would be curved on either end, but let's make the same approximation as we did then, anyway.

TIPS FOR SUCCESS

It is common in physics, especially at the early learning stages, for certain things to be *neglected*, as we see here. This is because it makes the concept clearer or the calculation easier. Practicing physicists use these kinds of short-cuts, as well. It works out because usually the thing being *neglected* is small enough that it does not significantly affect the answer. In the earlier example, the amount of time it takes the car to speed up and reach its cruising velocity is very small compared to the total time traveled.

Looking at this graph, and given what we learned, we can see that there are two distinct periods to the car's motion—the way to school and the way back. The average velocity for the drive to school is 0.5 km/minute. We can see that the average velocity for the drive back is −0.5 km/minute. If we plot the data showing velocity versus time, we get another graph (Figure 2.16):

Figure 2.16 Graph of velocity versus time for the drive to and from school.

We can learn a few things. First, we can derive a **v** versus *t* graph from a **d** versus *t* graph. Second, if we have a straight-line position–time graph that is positively or negatively sloped, it will yield a horizontal velocity graph. There are a few other interesting things to note. Just as we could use a position vs. time graph to determine velocity, we can use a velocity vs. time graph to determine position. We know that $\mathbf{v} = \mathbf{d}/t$. If we use a little algebra to re-arrange the equation, we see that $\mathbf{d} = \mathbf{v} \times t$. In Figure 2.16, we have velocity on the *y*-axis and time along the *x*-axis. Let's take just the first half of the motion. We get 0.5 km/minute × 10 minutes. The units for *minutes* cancel each other, and we get 5 km, which is the displacement for the trip to school. If we calculate the same for the return trip, we get −5 km. If we add them together, we see that the net displacement for the

whole trip is 0 km, which it should be because we started and ended at the same place.

TIPS FOR SUCCESS

You can treat units just like you treat numbers, so a km/km=1 (or, we say, it cancels out). This is good because it can tell us whether or not we have calculated everything with the correct units. For instance, if we end up with m × s for velocity instead of m/s, we know that something has gone wrong, and we need to check our math. This process is called dimensional analysis, and it is one of the best ways to check if your math makes sense in physics.

The area under a velocity curve represents the displacement. The velocity curve also tells us whether the car is speeding up. In our earlier example, we stated that the velocity was constant. So, the car is not speeding up. Graphically, you can see that the slope of these two lines is 0. This slope tells us that the car is not speeding up, or accelerating. We will do more with this information in a later chapter. For now, just remember that the area under the graph and the slope are the two important parts of the graph. Just like we could define a linear equation for the motion in a position vs. time graph, we can also define one for a velocity vs. time graph. As we said, the slope equals the acceleration, **a**. And in this graph, the *y*-intercept is $\mathbf{v}_0$. Thus, $\mathbf{v} = \mathbf{v}_0 + \mathbf{a}t$.

But what if the velocity is not constant? Let's look back at our jet-car example. At the beginning of the motion, as the car is speeding up, we saw that its position is a curve, as shown in Figure 2.17.

Figure 2.17 A graph is shown of the position of a jet-powered car during the time span when it is speeding up. The slope of a d vs. t graph is velocity. This is shown at two points. Instantaneous velocity at any point is the slope of the tangent at that point.

You do not have to do this, but you could, theoretically, take the instantaneous velocity at each point on this graph. If you did, you would get Figure 2.18, which is just a straight line with a positive slope.

Figure 2.18 The graph shows the velocity of a jet-powered car during the time span when it is speeding up.

Again, if we take the slope of the velocity vs. time graph, we get the acceleration, the rate of change of the velocity. And, if we take the area under the slope, we get back to the displacement.

Solving Problems using Velocity–Time Graphs

Most velocity vs. time graphs will be straight lines. When this is the case, our calculations are fairly simple.

WORKED EXAMPLE

Using Velocity Graph to Calculate Some Stuff: Jet Car

Use this figure to (a) find the displacement of the jet car over the time shown (b) calculate the rate of change (acceleration) of the velocity. (c) give the instantaneous velocity at 5 s, and (d) calculate the average velocity over the interval shown.

Strategy

a. The displacement is given by finding the area under the line in the velocity vs. time graph.
b. The acceleration is given by finding the slope of the velocity graph.
c. The instantaneous velocity can just be read off of the graph.
d. To find the average velocity, recall that $\mathbf{v}_{avg} = \frac{\Delta \mathbf{d}}{\Delta t} = \frac{\mathbf{d}_f - \mathbf{d}_0}{t_f - t_0}$

Solution

a. 1. Analyze the shape of the area to be calculated. In this case, the area is made up of a rectangle between 0 and 20 m/s stretching to 30 s. The area of a rectangle is length × width. Therefore, the area of this piece is 600 m.
   2. Above that is a triangle whose base is 30 s and height is 140 m/s. The area of a triangle is 0.5 × length × width. The area of this piece, therefore, is 2,100 m.
   3. Add them together to get a net displacement of 2,700 m.

b. 1. Take two points on the velocity line. Say, $t = 5$ s and $t = 25$ s. At $t = 5$ s, the value of $\mathbf{v} = 40$ m/s. At $t = 25$ s, $\mathbf{v} = 140$ m/s.

$$\mathbf{a} = \frac{\Delta \mathbf{v}}{\Delta t}$$

   2. Find the slope. $= \frac{100 \text{ m/s}}{20 \text{ s}}$

$$= 5 \text{ m/s}^2$$

c. The instantaneous velocity at $t = 5$ s, as we found in part (b) is just 40 m/s.

d. 1. Find the net displacement, which we found in part (a) was 2,700 m.
   2. Find the total time which for this case is 30 s.
   3. Divide 2,700 m/30 s = 90 m/s.

Discussion

The average velocity we calculated here makes sense if we look at the graph. 100m/s falls about halfway across the graph and since it is a straight line, we would expect about half the velocity to be above and half below.

TIPS FOR SUCCESS

You can have negative position, velocity, and acceleration on a graph that describes the way the object is moving. You should never see a graph with negative time on an axis. Why?

Most of the velocity vs. time graphs we will look at will be simple to interpret. Occasionally, we will look at curved graphs of velocity vs. time. More often, these curved graphs occur when something is speeding up, often from rest. Let's look back at a more realistic velocity vs. time graph of the jet car's motion that takes this *speeding up* stage into account.

Figure 2.19 The graph shows a more accurate graph of the velocity of a jet-powered car during the time span when it is speeding up.

WORKED EXAMPLE

Using Curvy Velocity Graph to Calculate Some Stuff: jet car, Take Two

Use Figure 2.19 to (a) find the approximate displacement of the jet car over the time shown, (b) calculate the instantaneous acceleration at $t = 30$ s, (c) find the instantaneous velocity at 30 s, and (d) calculate the approximate average velocity over the interval shown.

Strategy

a. Because this graph is an undefined curve, we have to estimate shapes over smaller intervals in order to find the areas.
b. Like when we were working with a curved displacement graph, we will need to take a tangent line at the instant we are interested and use that to calculate the instantaneous acceleration.
c. The instantaneous velocity can still be read off of the graph.
d. We will find the average velocity the same way we did in the previous example.

Solution

a.
1. This problem is more complicated than the last example. To get a good estimate, we should probably break the curve into four sections. 0 → 10 s, 10 → 20 s, 20 → 40 s, and 40 → 70 s.
2. Calculate the bottom rectangle (common to all pieces). 165 m/s × 70 s = 11,550 m.
3. Estimate a triangle at the top, and calculate the area for each section. Section 1 = 225 m; section 2 = 100 m + 450 m = 550 m; section 3 = 150 m + 1,300 m = 1,450 m; section 4 = 2,550 m.
4. Add them together to get a net displacement of 16,325 m.

b. Using the tangent line given, we find that the slope is 1 m/s^2.

c. The instantaneous velocity at $t = 30$ s, is 240 m/s.

d. 1. Find the net displacement, which we found in part (a), was 16,325 m.
   2. Find the total time, which for this case is 70 s.
   3. Divide $\frac{16{,}325 \text{ m}}{70 \text{ s}} \sim 233 \text{ m/s}$

Discussion

This is a much more complicated process than the first problem. If we were to use these estimates to come up with the average velocity over just the first 30 s we would get about 191 m/s. By approximating that curve with a line, we get an average velocity of 202.5 m/s. Depending on our purposes and how precise an answer we need, sometimes calling a curve a straight line is a worthwhile approximation.

Practice Problems

19.

Figure 2.20

Consider the velocity vs. time graph shown below of a person in an elevator. Suppose the elevator is initially at rest. It then speeds up for 3 seconds, maintains that velocity for 15 seconds, then slows down for 5 seconds until it stops. Find the instantaneous velocity at $t = 10$ s and $t = 23$ s.

a. Instantaneous velocity at $t = 10$ s and $t = 23$ s are 0 m/s and 0 m/s.
b. Instantaneous velocity at $t = 10$ s and $t = 23$ s are 0 m/s and 3 m/s.
c. Instantaneous velocity at $t = 10$ s and $t = 23$ s are 3 m/s and 0 m/s.
d. Instantaneous velocity at $t = 10$ s and $t = 23$ s are 3 m/s and 1.5 m/s.

20.

Figure 2.21

Calculate the net displacement and the average velocity of the elevator over the time interval shown.

a. Net displacement is 45 m and average velocity is 2.10 m/s.
b. Net displacement is 45 m and average velocity is 2.28 m/s.
c. Net displacement is 57 m and average velocity is 2.66 m/s.
d. Net displacement is 57 m and average velocity is 2.48 m/s.

Snap Lab

Graphing Motion, Take Two

In this activity, you will graph a moving ball's velocity vs. time.

- your graph from the earlier Graphing Motion Snap Lab!
- 1 piece of graph paper
- 1 pencil

Procedure

1. Take your graph from the earlier Graphing Motion Snap Lab! and use it to create a graph of velocity vs. time.
2. Use your graph to calculate the displacement.

GRASP CHECK

Describe the graph and explain what it means in terms of velocity and acceleration.

a. The graph shows a horizontal line indicating that the ball moved with a constant velocity, that is, it was not accelerating.
b. The graph shows a horizontal line indicating that the ball moved with a constant velocity, that is, it was accelerating.
c. The graph shows a horizontal line indicating that the ball moved with a variable velocity, that is, it was not accelerating.
d. The graph shows a horizontal line indicating that the ball moved with a variable velocity, that is, it was accelerating.

Check Your Understanding

21. What information could you obtain by looking at a velocity vs. time graph?

a. acceleration
b. direction of motion
c. reference frame of the motion

d. shortest path

22. How would you use a position vs. time graph to construct a velocity vs. time graph and vice versa?

a. Slope of position vs. time curve is used to construct velocity vs. time curve, and slope of velocity vs. time curve is used to construct position vs. time curve.
b. Slope of position vs. time curve is used to construct velocity vs. time curve, and area of velocity vs. time curve is used to construct position vs. time curve.
c. Area of position vs. time curve is used to construct velocity vs. time curve, and slope of velocity vs. time curve is used to construct position vs. time curve.
d. Area of position/time curve is used to construct velocity vs. time curve, and area of velocity vs. time curve is used to construct position vs. time curve.

KEY TERMS

acceleration the rate at which velocity changes
average speed distance traveled divided by time during which motion occurs
average velocity displacement divided by time over which displacement occurs
dependent variable the variable that changes as the independent variable changes
displacement the change in position of an object against a fixed axis
distance the length of the path actually traveled between an initial and a final position
independent variable the variable, usually along the horizontal axis of a graph, that does not change based on human or experimental action; in physics this is usually time
instantaneous speed speed at a specific instant in time
instantaneous velocity velocity at a specific instant in time
kinematics the study of motion without considering its causes
magnitude size or amount
position the location of an object at any particular time
reference frame a coordinate system from which the positions of objects are described
scalar a quantity that has magnitude but no direction
speed rate at which an object changes its location
tangent a line that touches another at exactly one point
vector a quantity that has both magnitude and direction
velocity the speed and direction of an object

SECTION SUMMARY

2.1 Relative Motion, Distance, and Displacement

- A description of motion depends on the reference frame from which it is described.
- The distance an object moves is the length of the path along which it moves.
- Displacement is the difference in the initial and final positions of an object.

2.2 Speed and Velocity

- Average speed is a scalar quantity that describes distance traveled divided by the time during which the motion occurs.
- Velocity is a vector quantity that describes the speed and direction of an object.
- Average velocity is displacement over the time period during which the displacement occurs. If the velocity is constant, then average velocity and instantaneous velocity are the same.

2.3 Position vs. Time Graphs

- Graphs can be used to analyze motion.
- The slope of a position vs. time graph is the velocity.
- For a straight line graph of position, the slope is the average velocity.
- To obtain the instantaneous velocity at a given moment for a curved graph, find the tangent line at that point and take its slope.

2.4 Velocity vs. Time Graphs

- The slope of a velocity vs. time graph is the acceleration.
- The area under a velocity vs. time curve is the displacement.
- Average velocity can be found in a velocity vs. time graph by taking the weighted average of all the velocities.

KEY EQUATIONS

2.1 Relative Motion, Distance, and Displacement

Displacement $\Delta \mathbf{d} = \mathbf{d_f} - \mathbf{d_0}$

2.2 Speed and Velocity

Average speed $v_{\text{avg}} = \frac{\textbf{distance}}{\textbf{time}}$

Average velocity $\mathbf{v}_{\text{avg}} = \frac{\Delta \mathbf{d}}{\Delta t} = \frac{\mathbf{d}_{\text{f}} - \mathbf{d}_0}{t_{\text{f}} - t_0}$

2.3 Position vs. Time Graphs

Displacement $\mathbf{d} = \mathbf{d}_0 + \mathbf{v}t$.

2.4 Velocity vs. Time Graphs

Velocity $\mathbf{v} = \mathbf{v}_0 + at$

Acceleration $\mathbf{a} = \frac{\Delta \mathbf{v}}{\Delta t}$

CHAPTER REVIEW

Concept Items

2.1 Relative Motion, Distance, and Displacement

1. Can one-dimensional motion have zero distance but a nonzero displacement? What about zero displacement but a nonzero distance?
   a. One-dimensional motion can have zero distance with a nonzero displacement. Displacement has both magnitude and direction, and it can also have zero displacement with nonzero distance because distance has only magnitude.
   b. One-dimensional motion can have zero distance with a nonzero displacement. Displacement has both magnitude and direction, but it cannot have zero displacement with nonzero distance because distance has only magnitude.
   c. One-dimensional motion cannot have zero distance with a nonzero displacement. Displacement has both magnitude and direction, but it can have zero displacement with nonzero distance because distance has only magnitude and any motion will be the distance it moves.
   d. One-dimensional motion cannot have zero distance with a nonzero displacement. Displacement has both magnitude and direction, and it cannot have zero displacement with nonzero distance because distance has only magnitude.

2. In which example would you be correct in describing an object in motion while your friend would also be correct in describing that same object as being at rest?
   a. You are driving a car toward the east and your friend drives past you in the opposite direction with the same speed. In your frame of reference, you will be in motion. In your friend's frame of reference, you will be at rest.
   b. You are driving a car toward the east and your friend is standing at the bus stop. In your frame of reference, you will be in motion. In your friend's frame of reference, you will be at rest.
   c. You are driving a car toward the east and your friend is standing at the bus stop. In your frame of reference, your friend will be moving toward the west. In your friend's frame of reference, he will be at rest.
   d. You are driving a car toward the east and your friend is standing at the bus stop. In your frame of reference, your friend will be moving toward the east. In your friend's frame of reference, he will be at rest.

3. What does your car's odometer record?
   a. displacement
   b. distance
   c. both distance and displacement
   d. the sum of distance and displacement

2.2 Speed and Velocity

4. In the definition of velocity, what physical quantity is changing over time?
   a. speed
   b. distance
   c. magnitude of displacement
   d. position vector

5. Which of the following best describes the relationship between instantaneous velocity and instantaneous speed?
   a. Both instantaneous speed and instantaneous velocity are the same, even when there is a change in direction.
   b. Instantaneous speed and instantaneous velocity cannot be the same even if there is no change in direction of motion.
   c. Magnitude of instantaneous velocity is equal to instantaneous speed.
   d. Magnitude of instantaneous velocity is always greater than instantaneous speed.

2.3 Position vs. Time Graphs

6. Use the graph to describe what the runner's motion looks like.

How are average velocity for only the first four seconds and instantaneous velocity related? What is the runner's net displacement over the time shown?
   a. The net displacement is 12 m and the average velocity is equal to the instantaneous velocity.
   b. The net displacement is 12 m and the average velocity

is two times the instantaneous velocity.

c. The net displacement is 10 + 12 = 22 m and the average velocity is equal to the instantaneous velocity.

d. The net displacement is 10 + 12 = 22 m and the average velocity is two times the instantaneous velocity.

7. A position vs. time graph of a frog swimming across a pond has two distinct straight-line sections. The slope of the first section is 1 m/s. The slope of the second section is 0 m/s. If each section lasts 1 second, then what is the frog's total average velocity?

a. 0 m/s
b. 2 m/s
c. 0.5 m/s
d. 1 m/s

2.4 Velocity vs. Time Graphs

8. A graph of velocity vs. time of a ship coming into a harbor is shown.

Describe the acceleration of the ship based on the graph.

a. The ship is moving in the forward direction at a steady rate. Then it accelerates in the forward direction and then decelerates.

b. The ship is moving in the forward direction at a steady rate. Then it turns around and starts decelerating, while traveling in the reverse direction. It then accelerates, but slowly.

c. The ship is moving in the forward direction at a steady rate. Then it decelerates in the forward direction, and then continues to slow down in the forward direction, but with more deceleration.

d. The ship is moving in the forward direction at a steady rate. Then it decelerates in the forward direction, and then continues to slow down in the forward direction, but with less deceleration.

Critical Thinking Items

2.1 Relative Motion, Distance, and Displacement

9. Boat A and Boat B are traveling at a constant speed in opposite directions when they pass each other. If a person in each boat describes motion based on the boat's own reference frame, will the description by a person in Boat A of Boat B's motion be the same as the description by a person in Boat B of Boat A's motion?

a. Yes, both persons will describe the same motion because motion is independent of the frame of reference.

b. Yes, both persons will describe the same motion because they will perceive the other as moving in the backward direction.

c. No, the motion described by each of them will be different because motion is a relative term.

d. No, the motion described by each of them will be different because the motion perceived by each will be opposite to each other.

10. Passenger A sits inside a moving train and throws a ball vertically upward. How would the motion of the ball be described by a fellow train passenger B and an observer C who is standing on the platform outside the train?

a. Passenger B sees that the ball has vertical, but no horizontal, motion. Observer C sees the ball has vertical as well as horizontal motion.

b. Passenger B sees the ball has vertical as well as horizontal motion. Observer C sees the ball has the vertical, but no horizontal, motion.

c. Passenger B sees the ball has horizontal but no vertical motion. Observer C sees the ball has vertical as well as horizontal motion.

d. Passenger B sees the ball has vertical as well as horizontal motion. Observer C sees the ball has horizontalbut no vertical motion.

2.2 Speed and Velocity

11. Is it possible to determine a car's instantaneous velocity from just the speedometer reading?

a. No, it reflects speed but not the direction.

b. No, it reflects the average speed of the car.

c. Yes, it sometimes reflects instantaneous velocity of the car.

d. Yes, it always reflects the instantaneous velocity of the car.

12. Terri, Aaron, and Jamal all walked along straight paths.

Terri walked 3.95 km north in 48 min. Aaron walked 2.65 km west in 31 min. Jamal walked 6.50 km south in 81 min. Which of the following correctly ranks the three boys in order from lowest to highest average speed?

a. Jamal, Terri, Aaron
b. Jamal, Aaron, Terri
c. Terri, Jamal, Aaron
d. Aaron, Terri, Jamal

13. Rhianna and Logan start at the same point and walk due north. Rhianna walks with an average velocity $v_{\text{avg,R}}$. Logan walks three times the distance in twice the time as Rhianna. Which of the following expresses Logan's average velocity in terms of $v_{\text{avg,R}}$?

a. Logan's average velocity = $1.5v_{avg,R}$.
b. Logan's average velocity = $\frac{2}{3}v_{avg,R}$.
c. Logan's average velocity = $3v_{avg,R}$.
d. Logan's average velocity = $\frac{1}{2}v_{avg,R}$.

2.3 Position vs. Time Graphs

14. Explain how you can use the graph of position vs. time to describe the change in velocity over time.

Identify the time (t_a, t_b, t_c, t_d, or t_e) at which at which the instantaneous velocity is greatest, the time at which it is zero, and the time at which it is negative.

2.4 Velocity vs. Time Graphs

15. Identify the time, or times, at which the instantaneous velocity is greatest, and the time, or times, at which it is negative. A sketch of velocity vs. time derived from the figure will aid in arriving at the correct answers.

a. The instantaneous velocity is greatest at *f*, and it is negative at *d*, *h*, *I*, *j*, and *k*.
b. The instantaneous velocity is greatest at *e*, and it is negative at *a*, *b*, and *f*.
c. The instantaneous velocity is greatest at *f*, and it is negative at *d*, *h*, *I*, *j*, and *k*
d. The instantaneous velocity is greatest at *d*, and it is negative at *a*, *b*, and *f*.

Problems

2.1 Relative Motion, Distance, and Displacement

16. In a coordinate system in which the direction to the right is positive, what are the distance and displacement of a person who walks 35 meters to the left, 18 meters to the right, and then 26 meters to the left?

a. Distance is 79 m and displacement is −43 m.
b. Distance is −79 m and displacement is 43 m.
c. Distance is 43 m and displacement is −79 m.
d. Distance is −43 m and displacement is 79 m.

17. Billy drops a ball from a height of 1 m. The ball bounces back to a height of 0.8 m, then bounces again to a height of 0.5 m, and bounces once more to a height of 0.2 m. Up is the positive direction. What are the total displacement of the ball and the total distance traveled by the ball?

a. The displacement is equal to -4 m and the distance is equal to 4 m.
b. The displacement is equal to 4 m and the distance is equal to 1 m.
c. The displacement is equal to 4 m and the distance is equal to 1 m.
d. The displacement is equal to -1 m and the distance is equal to 4 m.

2.2 Speed and Velocity

18. You sit in a car that is moving at an average speed of 86.4 km/h. During the 3.3 s that you glance out the window, how far has the car traveled?

a. 7.27 m
b. 79 m
c. 285 km
d. 1026 m

2.3 Position vs. Time Graphs

19. Using the graph, what is the average velocity for the whole 10 seconds?

a. The total average velocity is 0 m/s.
b. The total average velocity is 1.2 m/s.
c. The total average velocity is 1.5 m/s.
d. The total average velocity is 3.0 m/s.

20. A train starts from rest and speeds up for 15 minutes until it reaches a constant velocity of 100 miles/hour. It stays at this speed for half an hour. Then it slows down for another 15 minutes until it is still. Which of the following correctly describes the position vs time graph of the train's journey?

a. The first 15 minutes is a curve that is concave upward, the middle portion is a straight line with slope 100 miles/hour, and the last portion is a concave downward curve.
b. The first 15 minutes is a curve that is concave downward, the middle portion is a straight line with slope 100 miles/hour, and the last portion is a concave upward curve.
c. The first 15 minutes is a curve that is concave upward, the middle portion is a straight line with slope zero, and the last portion is a concave downward curve.
d. The first 15 minutes is a curve that is concave downward, the middle portion is a straight line with slope zero, and the last portion is a concave upward curve.

2.4 Velocity vs. Time Graphs

21. You are characterizing the motion of an object by measuring the location of the object at discrete moments in time. What is the minimum number of data points you would need to estimate the average acceleration of the object?

a. 1
b. 2
c. 3
d. 4

22. Which option best describes the average acceleration from 40 to 70 s?

a. It is negative and smaller in magnitude than the initial acceleration.
b. It is negative and larger in magnitude than the initial acceleration.
c. It is positive and smaller in magnitude than the initial acceleration.
d. It is positive and larger in magnitude than the initial acceleration.

23. The graph shows velocity vs. time.

Calculate the net displacement using seven different divisions. Calculate it again using two divisions: 0 → 40 s

and 40 → 70 s . Compare. Using both, calculate the average velocity.

a. Displacement and average velocity using seven divisions are 14,312.5 m and 204.5 m/s while with two divisions are 15,500 m and 221.4 m/s respectively.
b. Displacement and average velocity using seven divisions are 15,500 m and 221.4 m/s while with two divisions are 14,312.5 m and 204.5 m/s respectively.
c. Displacement and average velocity using seven divisions are 15,500 m and 204.5 m/s while with two divisions are 14,312.5 m and 221.4 m/s respectively.
d. Displacement and average velocity using seven divisions are 14,312.5 m and 221.4 m/s while with two divisions are 15,500 m and 204.5 m/s respectively.

Performance Task

2.4 Velocity vs. Time Graphs

24. The National Mall in Washington, DC, is a national park containing most of the highly treasured memorials and museums of the United States. However, the National Mall also hosts many events and concerts. The map in shows the area for a benefit concert during which the president will speak. The concert stage is near the Lincoln Memorial. The seats and standing room for the crowd will stretch from the stage east to near the Washington Monument, as shown on the map. You are planning the logistics for the concert. Use the map scale to measure any distances needed to make the calculations below.

The park has three new long-distance speakers. They would like to use these speakers to broadcast the concert audio to other parts of the National Mall. The speakers can project sound up to 0.35 miles away but they must be connected to one of the power supplies within the concert area. What is the minimum amount of wire needed for each speaker, in miles, in order to project the audio to the following areas? Assume that wire cannot be placed over buildings or any memorials.

Part A. The center of the Jefferson Memorial using power supply 1 (This will involve an elevated wire that can travel over water.)

Part B. The center of the Ellipse using power supply 3 (This wire cannot travel over water.)

Part C. The president's motorcade will be traveling to the concert from the White House. To avoid concert traffic, the motorcade travels from the White House west down E Street and then turns south on 23rd Street to reach the Lincoln memorial. What minimum speed, in miles per hour to the nearest tenth, would the motorcade have to travel to make the trip in 5 minutes?

Part D. The president could also simply fly from the White House to the Lincoln Memorial using the presidential helicopter, Marine 1. How long would it take Marine 1, traveling slowly at 30 mph, to travel from directly above the White House landing zone (LZ) to directly above the Lincoln Memorial LZ? Disregard liftoff and landing times and report the travel time in minutes to the nearest minute.

TEST PREP

Multiple Choice

2.1 Relative Motion, Distance, and Displacement

25. Why should you specify a reference frame when describing motion?

a. a description of motion depends on the reference frame
b. motion appears the same in all reference frames
c. reference frames affect the motion of an object
d. you can see motion better from certain reference frames

26. Which of the following is true for the displacement of an object?

a. It is always equal to the distance the object moved

between its initial and final positions.
b. It is both the straight line distance the object moved as well as the direction of its motion.
c. It is the direction the object moved between its initial and final positions.
d. It is the straight line distance the object moved between its initial and final positions.

27. If a biker rides west for 50 miles from his starting position, then turns and bikes back east 80 miles. What is his net displacement?
a. 130 miles
b. 30 miles east
c. 30 miles west
d. Cannot be determined from the information given

28. Suppose a train is moving along a track. Is there a single, correct reference frame from which to describe the train's motion?
a. Yes, there is a single, correct frame of reference because motion is a relative term.
b. Yes, there is a single, correct frame of reference which is in terms of Earth's position.
c. No, there is not a single, correct frame of reference because motion is a relative term.
d. No, there is not a single, correct frame of reference because motion is independent of frame of reference.

29. If a space shuttle orbits Earth once, what is the shuttle's distance traveled and displacement?
a. Distance and displacement both are zero.
b. Distance is circumference of the circular orbit while displacement is zero.
c. Distance is zero while the displacement is circumference of the circular orbit.
d. Distance and displacement both are equal to circumference of the circular orbit.

2.2 Speed and Velocity

30. Four bicyclists travel different distances and times along a straight path. Which cyclist traveled with the greatest average speed?
a. Cyclist 1 travels 95 m in 27 s.
b. Cyclist 2 travels 87 m in 22 s.
c. Cyclist 3 travels 106 m in 26 s.
d. Cyclist 4 travels 108 m in 24 s.

31. A car travels with an average velocity of 23 m/s for 82 s. Which of the following could NOT have been the car's displacement?
a. 1,700 m east
b. 1,900 m west
c. 1,600 m north
d. 1,500 m south

32. A bicyclist covers the first leg of a journey that is d_1 meters long in t_1 seconds, at a speed of v_1 m/s, and the second leg of d_2 meters in t_2 seconds, at a speed of v_2 m/s. If his average speed is equal to the average of v_1 and v_2, then which of the following is true?
a. $t_1 = t_2$
b. $t_1 \neq t_2$
c. $d_1 = d_2$
d. $d_1 \neq d_2$

33. A car is moving on a straight road at a constant speed in a single direction. Which of the following statements is true?
a. Average velocity is zero.
b. The magnitude of average velocity is equal to the average speed.
c. The magnitude of average velocity is greater than the average speed.
d. The magnitude of average velocity is less than the average speed.

2.3 Position vs. Time Graphs

34. What is the slope of a straight line graph of position vs. time?
a. Velocity
b. Displacement
c. Distance
d. Acceleration

35. Using the graph, what is the runner's velocity from 4 to 10 s?

a. −3 m/s
b. 0 m/s
c. 1.2 m/s
d. 3 m/s

2.4 Velocity vs. Time Graphs

36. What does the area under a velocity vs. time graph line represent?
a. acceleration
b. displacement
c. distance
d. instantaneous velocity

37. An object is moving along a straight path with constant acceleration. A velocity vs. time graph starts at 0 and ends at $10\ \text{m/s}$, stretching over a time-span of $15\ \text{s}$. What is the object's net displacement?
a. $75\ \text{m}$
b. $130\ \text{m}$
c. $150\ \text{m}$
d. cannot be determined from the information given

Short Answer

2.1 Relative Motion, Distance, and Displacement

38. While standing on a sidewalk facing the road, you see a bicyclist passing by toward your right. In the reference frame of the bicyclist, in which direction are you moving?
a. in the same direction of motion as the bicyclist
b. in the direction opposite the motion of the bicyclist
c. stationary with respect to the bicyclist
d. in the direction of velocity of the bicyclist

39. Maud sends her bowling ball straight down the center of the lane, getting a strike. The ball is brought back to the holder mechanically. What are the ball's net displacement and distance traveled?
a. Displacement of the ball is twice the length of the lane, while the distance is zero.
b. Displacement of the ball is zero, while the distance is twice the length of the lane.
c. Both the displacement and distance for the ball are equal to zero.
d. Both the displacement and distance for the ball are twice the length of the lane.

40. A fly buzzes four and a half times around Kit Yan's head. The fly ends up on the opposite side from where it started. If the diameter of his head is $14\ \text{cm}$, what is the total distance the fly travels and its total displacement?
a. The distance is $63\pi\ \text{cm}$ with a displacement of zero.
b. The distance is $7\ \text{cm}$ with a displacement of zero.
c. The distance is $63\pi\ \text{cm}$ with a displacement of $14\ \text{cm}$.
d. The distance is $7\ \text{cm}$ with a displacement of $63\pi\ \text{cm}$.

2.2 Speed and Velocity

41. Rob drove to the nearest hospital with an average speed of v m/s in t seconds. In terms of t, if he drives home on the same path, but with an average speed of 3v m/s, how long is the return trip home?
a. t/6
b. t/3
c. 3t
d. 6t

42. What can you infer from the statement, *Velocity of an object is zero*?
a. Object is in linear motion with constant velocity.
b. Object is moving at a constant speed.
c. Object is either at rest or it returns to the initial point.
d. Object is moving in a straight line without changing its direction.

43. An object has an average speed of 7.4 km/h. Which of the following describes two ways you could increase the average speed of the object to 14.8 km/h?
a. Reduce the distance that the object travels by half, keeping the time constant, or keep the distance constant and double the time.
b. Double the distance that the object travels, keeping the time constant, or keep the distance constant and reduce the time by half.
c. Reduce the distance that the object travels to one-fourth, keeping the time constant, or keep the distance constant and increase the time by fourfold.
d. Increase the distance by fourfold, keeping the time constant, or keep the distance constant and reduce the time by one-fourth.

44. Swimming one lap in a pool is defined as going across a pool and back again. If a swimmer swims 3 laps in 9 minutes, how can his average velocity be zero?
a. His average velocity is zero because his total distance is zero.
b. His average velocity is zero because his total displacement is zero.
c. His average velocity is zero because the number of laps completed is an odd number.
d. His average velocity is zero because the velocity of each successive lap is equal and opposite.

2.3 Position vs. Time Graphs

45. A hockey puck is shot down the arena in a straight line. Assume it does not slow until it is stopped by an opposing player who sends it back in the direction it came. The players are 20 m apart and it takes 1 s for the puck to go there and back. Which of the following describes the graph of the displacement over time? Consider the initial direction of the puck to be positive.

a. The graph is an upward opening V.
b. The graph is a downward opening V.
c. The graph is an upward opening U.
d. The graph is downward opening U.

46. A defensive player kicks a soccer ball 20 m back to her own goalie. It stops just as it reaches her. She sends it back to the player. Without knowing the time it takes, draw a rough sketch of the displacement over time. Does this graph look similar to the graph of the hockey puck from the previous question?

a. Yes, the graph is similar to the graph of the hockey puck.
b. No, the graph is not similar to the graph of the hockey puck.
c. The graphs cannot be compared without knowing the time the soccer ball was rolling.

47. What are the net displacement, total distance traveled, and total average velocity in the previous two problems?

a. net displacement = 0 m, total distance = 20 m, total average velocity = 20 m/s
b. net displacement = 0 m, total distance = 40 m, total average velocity = 20 m/s
c. net displacement = 0 m, total distance = 20 m, total average velocity = 0 m/s
d. net displacement = 0 m, total distance = 40 m, total average velocity = 0 m/s

48. A bee flies straight at someone and then back to its hive along the same path. Assuming it takes no time for the bee to speed up or slow down, except at the moment it changes direction, how would the graph of position vs time look? Consider the initial direction to be positive.

a. The graph will look like a downward opening V shape.
b. The graph will look like an upward opening V shape.
c. The graph will look like a downward opening parabola.
d. The graph will look like an upward opening parabola.

2.4 Velocity vs. Time Graphs

49. What would the velocity vs. time graph of the object whose position is shown in the graph look like?

a. It is a straight line with negative slope.
b. It is a straight line with positive slope.
c. It is a horizontal line at some negative value.
d. It is a horizontal line at some positive value.

50. Which statement correctly describes the object's speed, as well as what a graph of acceleration vs. time would look like?

a. The object is not speeding up, and the acceleration vs. time graph is a horizontal line at some negative value.
b. The object is not speeding up, and the acceleration vs. time graph is a horizontal line at some positive value.
c. The object is speeding up, and the acceleration vs. time graph is a horizontal line at some negative value.
d. The object is speeding up, and the acceleration vs. time graph is a horizontal line at some positive value.

51. Calculate that object's net displacement over the time shown.

a. 540 m
b. 2,520 m
c. 2,790 m
d. 5,040 m

a. 18 m/s
b. 84 m/s
c. 93 m/s
d. 168 m/s

52. What is the object's average velocity?

Extended Response

2.1 Relative Motion, Distance, and Displacement

53. Find the distance traveled from the starting point for each path.

Which path has the maximum distance?

a. The distance for Path A is 6 m, Path B is 4 m, Path C is 12 m and for Path D is 7 m. The net displacement for Path A is 7 m, Path B is −4m, Path C is 8 m and for Path D is −5m. Path C has maximum distance and it is equal to 12 meters.
b. The distance for Path A is 6 m, Path B is 4 m, Path C is 8 m and for Path D is 7 m. The net displacement for Path A is 6 m, Path B is −4m, Path C is 12 m and for Path D is −5 m. Path A has maximum distance and it is equal to 6 meters.
c. The distance for Path A is 6 m, Path B is 4 m, Path C is 12 m and for Path D is 7 m. The net displacement for Path A is 6 m, Path B is −4 m, Path C is 8 m and for Path D is −5 m. Path C has maximum distance and it is equal to 12 meters.
d. The distance for Path A is 6 m, Path B is −4 m, Path C is 12 m and for Path D is −5 m. The net displacement for Path A is 7 m, Path B is 4 m, Path C is 8 m and for Path D is 7 m. Path A has maximum distance and it is equal to 6 m.

54. Alan starts from his home and walks 1.3 km east to the library. He walks an additional 0.68 km east to a music store. From there, he walks 1.1 km north to a friend's house and an additional 0.42 km north to a grocery store before he finally returns home along the same path. What is his final displacement and total distance traveled?

a. Displacement is 0 km and distance is 7 km.
b. Displacement is 0 km and distance is 3.5 km.
c. Displacement is 7 km towards west and distance is 7 km.
d. Displacement is 3.5 km towards east and distance is 3.5 km.

2.2 Speed and Velocity

55. Two runners start at the same point and jog at a constant speed along a straight path. Runner A starts at time t = 0 s, and Runner B starts at time t = 2.5 s. The runners both reach a distance 64 m from the starting point at time t = 25 s. If the runners continue at the same speeds, how far from the starting point will each be at time t = 45 s?

a. Runner A will be 72×10^3 m away and Runner B

will be 59.5×10^3 m away from the starting point.

b. Runner A will be 1.2×10^2 m away and runner B will be 1.1×10^2 m away from the starting point.

c. Runner A will be 1.2×10^2 m away and Runner B will be 1.3×10^2 m away from the starting point.

d. Runner A will be 7.2×10^2 m away and Runner B will be 1.3×10^2 m away from the starting point.

56. A father and his daughter go to the bus stop that is located 75 m from their front door. The father walks in a straight line while his daughter runs along a varied path. Despite the different paths, they both end up at the bus stop at the same time. The father's average speed is 2.2 m/s, and his daughter's average speed is 3.5 m/s. (a) How long does it take the father and daughter to reach the bus stop? (b) What was the daughter's total distance traveled? (c) If the daughter maintained her same average speed and traveled in a straight line like her father, how far beyond the bus stop would she have traveled?

a. (a) 21.43 s (b) 75 m (c) 0 m

b. (a) 21.43 s (b) 119 m (c) 44 m

c. (a) 34 s (b) 75 m (c) 0 m

d. (a) 34 s (b) 119 m (c) 44 m

2.3 Position vs. Time Graphs

57. What kind of motion would create a position graph like the one shown?

a. uniform motion

b. any motion that accelerates

c. motion that stops and then starts

d. motion that has constant velocity

58. What is the average velocity for the whole time period shown in the graph?

a. $-\frac{1}{3}$ m/s

b. $-\frac{3}{4}$ m/s

c. $\frac{1}{3}$ m/s

d. $\frac{3}{4}$ m/s

2.4 Velocity vs. Time Graphs

59. Consider the motion of the object whose velocity is charted in the graph.

During which points is the object slowing down and speeding up?

a. It is slowing down between *d* and *e*. It is speeding up between *a* and *d* and *e* and *h*

b. It is slowing down between *a* and *d* and *e* and *h*. It is speeding up between *d* and *e* and then after *i*.

c. It is slowing down between *d* and *e* and then after *h*. It is speeding up between *a* and *d* and *e* and *h*.

d. It is slowing down between *a* and *d* and *e* and *h*. It is speeding up between *d* and *e* and then after *i*.

60. Divide the graph into approximate sections, and use those sections to graph the velocity vs. time of the object.

Then calculate the acceleration during each section, and calculate the approximate average velocity.

a. Acceleration is zero and average velocity is 1.25 m/s.
b. Acceleration is constant with some positive value and average velocity is 1.25 m/s.
c. Acceleration is zero and average velocity is 0.25 m/s.
d. Acceleration is constant with some positive value and average velocity is 0.25 m/s.

CHAPTER 3
Acceleration

Figure 3.1 A plane slows down as it comes in for landing in St. Maarten. Its acceleration is in the opposite direction of its velocity. (Steve Conry, Flickr)

Chapter Outline

3.1 Acceleration

3.2 Representing Acceleration with Equations and Graphs

INTRODUCTION You may have heard the term *accelerator*, referring to the gas pedal in a car. When the gas pedal is pushed down, the flow of gasoline to the engine increases, which increases the car's velocity. Pushing on the gas pedal results in acceleration because the velocity of the car increases, and acceleration is defined as a change in velocity. You need two quantities to define velocity: a speed and a direction. Changing either of these quantities, or both together, changes the velocity. You may be surprised to learn that pushing on the brake pedal or turning the steering wheel also causes acceleration. The first reduces the *speed* and so changes the velocity, and the second changes the *direction* and also changes the velocity.

In fact, any change in velocity—whether positive, negative, directional, or any combination of these—is called an acceleration in physics. The plane in the picture is said to be accelerating because its velocity is decreasing as it prepares to land. To begin our study of acceleration, we need to have a clear understanding of what acceleration means.

3.1 Acceleration

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Explain acceleration and determine the direction and magnitude of acceleration in one dimension
- Analyze motion in one dimension using kinematic equations and graphic representations

Section Key Terms

average acceleration	instantaneous acceleration	negative acceleration

Defining Acceleration

Throughout this chapter we will use the following terms: *time*, *displacement*, *velocity*, and *acceleration*. Recall that each of these terms has a designated variable and SI unit of measurement as follows:

- Time: t, measured in seconds (s)
- Displacement: Δd, measured in meters (m)
- Velocity: v, measured in meters per second (m/s)
- Acceleration: a, measured in meters per second per second (m/s^2, also called meters per second squared)
- Also note the following:
  - Δ means *change in*
  - The subscript 0 refers to an initial value; sometimes subscript i is instead used to refer to initial value.
  - The subscript f refers to final value
  - A bar over a symbol, such as $\overline{a}$, means *average*

Acceleration is the change in velocity divided by a period of time during which the change occurs. The SI units of velocity are m/s and the SI units for time are s, so the SI units for acceleration are m/s^2. **Average acceleration** is given by

$$\overline{a} = \frac{\Delta v}{\Delta t} = \frac{v_f - v_0}{t_f - t_0}.$$

Average acceleration is distinguished from **instantaneous acceleration**, which is acceleration at a specific instant in time. The magnitude of acceleration is often not constant over time. For example, runners in a race accelerate at a greater rate in the first second of a race than during the following seconds. You do not need to know all the instantaneous accelerations at all times to calculate average acceleration. All you need to know is the change in velocity (i.e., the final velocity minus the initial velocity) and the change in time (i.e., the final time minus the initial time), as shown in the formula. Note that the average acceleration can be positive, negative, or zero. A **negative acceleration** is simply an acceleration in the negative direction.

Keep in mind that although acceleration points in the same direction as the *change* in velocity, it is not always in the direction of the velocity itself. When an object slows down, its acceleration is opposite to the direction of its velocity. In everyday language, this is called deceleration; but in physics, it is acceleration—whose direction happens to be opposite that of the velocity. For now, let us assume that motion to the right along the x-axis is *positive* and motion to the left is *negative*.

Figure 3.2 shows a car with positive acceleration in (a) and negative acceleration in (b). The arrows represent vectors showing both direction and magnitude of velocity and acceleration.

Figure 3.2 The car is speeding up in (a) and slowing down in (b).

Velocity and acceleration are both vector quantities. Recall that vectors have both magnitude and direction. An object traveling at a constant velocity—therefore having no acceleration—does accelerate if it changes direction. So, turning the steering wheel of a moving car makes the car accelerate because the velocity changes direction.

Virtual Physics

The Moving Man

With this animation in , you can produce both variations of acceleration and velocity shown in Figure 3.2, plus a few more variations. Vary the velocity and acceleration by sliding the red and green markers along the scales. Keeping the velocity marker near zero will make the effect of acceleration more obvious. Try changing acceleration from positive to negative while the man is moving. We will come back to this animation and look at the *Charts* view when we study graphical representation of motion.

Click to view content (https://archive.cnx.org/specials/e2ca52af-8c6b-450e-ac2f-9300b38e8739/moving-man/)

GRASP CHECK

Figure 3.3

Which part, (a) or (b), is represented when the velocity vector is on the positive side of the scale and the acceleration vector is set on the negative side of the scale? What does the car's motion look like for the given scenario?

a. Part (a). The car is slowing down because the acceleration and the velocity vectors are acting in the opposite direction.
b. Part (a). The car is speeding up because the acceleration and the velocity vectors are acting in the same direction.
c. Part (b). The car is slowing down because the acceleration and velocity vectors are acting in the opposite directions.
d. Part (b). The car is speeding up because the acceleration and the velocity vectors are acting in the same direction.

Calculating Average Acceleration

Look back at the equation for average acceleration. You can see that the calculation of average acceleration involves three values: change in time, (Δt); change in velocity, (Δv); and acceleration (a).

Change in time is often stated as a time interval, and change in velocity can often be calculated by subtracting the initial velocity from the final velocity. Average acceleration is then simply change in velocity divided by change in time. Before you begin calculating, be sure that all distances and times have been converted to meters and seconds. Look at these examples of acceleration of a subway train.

WORKED EXAMPLE

An Accelerating Subway Train

A subway train accelerates from rest to 30.0 km/h in 20.0 s. What is the average acceleration during that time interval?

Strategy

Start by making a simple sketch.

Figure 3.4

This problem involves four steps:

1. Convert to units of meters and seconds.
2. Determine the change in velocity.
3. Determine the change in time.
4. Use these values to calculate the average acceleration.

Solution

1. Identify the knowns. Be sure to read the problem for given information, which may not *look* like numbers. When the problem states that the train starts from rest, you can write down that the initial velocity is 0 m/s. Therefore, $v_0 = 0$; $v_f = 30.0$ km/h; and $\Delta t = 20.0$ s.
2. Convert the units.

$$\frac{30.0\text{ km}}{\text{h}} \times \frac{10^3\text{ m}}{1\text{ km}} \times \frac{1\text{ h}}{3600\text{ s}} = 8.333\ \frac{\text{m}}{\text{s}} \tag{3.1}$$

3. Calculate change in velocity, $\Delta v = v_f - v_0 = 8.333\text{ m/s} - 0 = +8.333\text{ m/s}$, where the plus sign means the change in velocity is to the right.
4. We know Δt, so all we have to do is insert the known values into the formula for average acceleration.

$$\bar{a} = \frac{\Delta v}{\Delta t} = \frac{8.333\text{ m/s}}{20.00\text{ s}} = +0.417\frac{\text{m}}{\text{s}^2} \tag{3.2}$$

Discussion

The plus sign in the answer means that acceleration is to the right. This is a reasonable conclusion because the train starts from rest and ends up with a velocity directed to the right (i.e., positive). So, acceleration is in the same direction as the *change* in velocity, as it should be.

WORKED EXAMPLE

An Accelerating Subway Train

Now, suppose that at the end of its trip, the train slows to a stop in 8.00 s from a speed of 30.0 km/h. What is its average acceleration during this time?

Strategy

Again, make a simple sketch.

Figure 3.5

In this case, the train is decelerating and its acceleration is negative because it is pointing to the left. As in the previous example, we must find the change in velocity and change in time, then solve for acceleration.

Solution

1. Identify the knowns: $v_0 = 30.0$ km/h; $v_f = 0$; and $\Delta t = 8.00$ s.
2. Convert the units. From the first problem, we know that 30.0 km/h = 8.333 m/s.
3. Calculate change in velocity, $\Delta v = v_f - v_0 = 0 - 8.333 \text{ m/s} = -8.333 \text{ m/s}$, where the minus sign means that the change in velocity points to the left.
4. We know $\Delta t = 8.00$ s, so all we have to do is insert the known values into the equation for average acceleration.

$$\bar{a} = \frac{\Delta v}{\Delta t} = \frac{-8.333 \text{ m/s}}{8.00 \text{ s}} = -1.04 \frac{\text{m}}{\text{s}^2} \tag{3.3}$$

Discussion

The minus sign indicates that acceleration is to the left. This is reasonable because the train initially has a positive velocity in this problem, and a negative acceleration would reduce the velocity. Again, acceleration is in the same direction as the *change* in velocity, which is negative in this case. This acceleration can be called a deceleration because it has a direction opposite to the velocity.

TIPS FOR SUCCESS

- It is easier to get plus and minus signs correct if you always assume that motion is away from zero and toward positive values on the *x*-axis. This way *v* always starts off being positive and points to the right. If speed is increasing, then acceleration is positive and also points to the right. If speed is decreasing, then acceleration is negative and points to the left.
- It is a good idea to carry two extra significant figures from step-to-step when making calculations. Do not round off with each step. When you arrive at the final answer, apply the rules of significant figures for the operations you carried out and round to the correct number of digits. Sometimes this will make your answer slightly more accurate.

Practice Problems

1. A cheetah can accelerate from rest to a speed of 30.0 m/s in 7.00 s. What is its acceleration?
   a. -0.23 m/s^2
   b. -4.29 m/s^2
   c. 0.23 m/s^2
   d. 4.29 m/s^2

2. A women backs her car out of her garage with an acceleration of 1.40 m/s^2. How long does it take her to reach a speed of 2.00 m/s?
   a. 0.70 s
   b. 1.43 s
   c. 2.80 s
   d. 3.40 s

WATCH PHYSICS

Acceleration

This video shows the basic calculation of acceleration and some useful unit conversions.

Click to view content (https://www.khanacademy.org/embed_video?v=FOkQszg1-j8)

GRASP CHECK

Why is acceleration a vector quantity?

a. It is a vector quantity because it has magnitude as well as direction.
b. It is a vector quantity because it has magnitude but no direction.
c. It is a vector quantity because it is calculated from distance and time.
d. It is a vector quantity because it is calculated from speed and time.

GRASP CHECK

What will be the change in velocity each second if acceleration is 10 m/s/s?

a. An acceleration of $10\ \text{m/s/s}$ means that every second, the velocity increases by $10\ \text{m/s}$.
b. An acceleration of $10\ \text{m/s/s}$ means that every second, the velocity decreases by $10\ \text{m/s}$.
c. An acceleration of $10\ \text{m/s/s}$ means that every $10\ \text{seconds}$, the velocity increases by $10\ \text{m/s}$.
d. An acceleration of $10\ \text{m/s/s}$ means that every $10\ \text{seconds}$, the velocity decreases by $10\ \text{m/s}$.

Snap Lab

Measure the Acceleration of a Bicycle on a Slope

In this lab you will take measurements to determine if the acceleration of a moving bicycle is constant. If the acceleration is constant, then the following relationships hold: $\overline{v} = \frac{\Delta d}{\Delta t} = \frac{v_0 + v_f}{2}$ If $v_0 = 0$, then $v_f = 2\overline{v}$ and $\overline{a} = \frac{v_f}{\Delta t}$

You will work in pairs to measure and record data for a bicycle coasting down an incline on a smooth, gentle slope. The data will consist of distances traveled and elapsed times.

- Find an open area to minimize the risk of injury during this lab.

- stopwatch
- measuring tape
- bicycle

1. Find a gentle, paved slope, such as an incline on a bike path. The more gentle the slope, the more accurate your data will likely be.
2. Mark uniform distances along the slope, such as 5 m, 10 m, etc.
3. Determine the following roles: the bike rider, the timer, and the recorder. The recorder should create a data table to collect the distance and time data.
4. Have the rider at the starting point at rest on the bike. When the timer calls *Start*, the timer starts the stopwatch and the rider begins coasting down the slope on the bike without pedaling.
5. Have the timer call out the elapsed times as the bike passes each marked point. The recorder should record the times in the data table. It may be necessary to repeat the process to practice roles and make necessary adjustments.
6. Once acceptable data has been recorded, switch roles. Repeat Steps 3–5 to collect a second set of data.
7. Switch roles again to collect a third set of data.
8. Calculate average acceleration for each set of distance-time data. If your result for $\overline{a}$ is not the same for different pairs of Δv and Δt, then acceleration is not constant.
9. Interpret your results.

GRASP CHECK

If you graph the average velocity (y-axis) vs. the elapsed time (x-axis), what would the graph look like if acceleration is uniform?

a. a horizontal line on the graph
b. a diagonal line on the graph
c. an upward-facing parabola on the graph
d. a downward-facing parabola on the graph

Check Your Understanding

3. What are three ways an object can accelerate?
   a. By speeding up, maintaining constant velocity, or changing direction
   b. By speeding up, slowing down, or changing direction
   c. By maintaining constant velocity, slowing down, or changing direction
   d. By speeding up, slowing down, or maintaining constant velocity

4. What is the difference between average acceleration and instantaneous acceleration?
   a. Average acceleration is the change in displacement divided by the elapsed time; instantaneous acceleration is the acceleration at a given point in time.
   b. Average acceleration is acceleration at a given point in time; instantaneous acceleration is the change in displacement divided by the elapsed time.
   c. Average acceleration is the change in velocity divided by the elapsed time; instantaneous acceleration is acceleration at a given point in time.
   d. Average acceleration is acceleration at a given point in time; instantaneous acceleration is the change in velocity divided by the elapsed time.

5. What is the rate of change of velocity called?
   a. Time
   b. Displacement
   c. Velocity
   d. Acceleration

3.2 Representing Acceleration with Equations and Graphs

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Explain the kinematic equations related to acceleration and illustrate them with graphs
- Apply the kinematic equations and related graphs to problems involving acceleration

Section Key Terms

acceleration due to gravity	kinematic equations	uniform acceleration

How the Kinematic Equations are Related to Acceleration

We are studying concepts related to motion: time, displacement, velocity, and especially acceleration. We are only concerned with motion in one dimension. The **kinematic equations** apply to conditions of constant acceleration and show how these concepts are related. **Constant acceleration** is acceleration that does not change over time. The first kinematic equation relates displacement d, average velocity $\overline{v}$, and time t.

$$d = d_0 + \overline{v}t \tag{3.4}$$

The initial displacement d_0 is often 0, in which case the equation can be written as $\overline{v} = \frac{d}{t}$

This equation says that average velocity is displacement per unit time. We will express velocity in meters per second. If we graph displacement versus time, as in Figure 3.6, the slope will be the velocity. Whenever a rate, such as velocity, is represented graphically, time is usually taken to be the independent variable and is plotted along the *x* axis.

Figure 3.6 The slope of displacement versus time is velocity.

The second kinematic equation, another expression for average velocity $\overline{v}$, is simply the initial velocity plus the final velocity divided by two.

$$\overline{v} = \frac{v_0 + v_f}{2} \tag{3.5}$$

Now we come to our main focus of this chapter; namely, the kinematic equations that describe motion with constant acceleration. In the third kinematic equation, acceleration is the rate at which velocity increases, so velocity at any point equals initial velocity plus acceleration multiplied by time

$$v = v_0 + at \text{ Also, if we start from rest } (v_0 = 0\ ), \text{ we can write } a = \frac{v}{t} \tag{3.6}$$

Note that this third kinematic equation does not have displacement in it. Therefore, if you do not know the displacement and are not trying to solve for a displacement, this equation might be a good one to use.

The third kinematic equation is also represented by the graph in Figure 3.7.

Figure 3.7 The slope of velocity versus time is acceleration.

The fourth kinematic equation shows how displacement is related to acceleration

$$d = d_0 + v_0 t + \frac{1}{2}at^2. \tag{3.7}$$

When starting at the origin, $d_0 = 0$ and, when starting from rest, $v_0 = 0$, in which case the equation can be written as

$$a = \frac{2d}{t^2}.$$

This equation tells us that, for constant acceleration, the slope of a plot of $2d$ versus t^2 is acceleration, as shown in Figure 3.8.

Figure 3.8 When acceleration is constant, the slope of $2d$ versus t^2 gives the acceleration.

The fifth kinematic equation relates velocity, acceleration, and displacement

$$v^2 = v_0^2 + 2a(d - d_0). \tag{3.8}$$

This equation is useful for when we do not know, or do not need to know, the time.

When starting from rest, the fifth equation simplifies to

$$a = \frac{v^2}{2d}.$$

According to this equation, a graph of velocity squared versus twice the displacement will have a slope equal to acceleration.

Figure 3.9

Note that, in reality, knowns and unknowns will vary. Sometimes you will want to rearrange a kinematic equation so that the knowns are the values on the axes and the unknown is the slope. Sometimes the intercept will not be at the origin (0,0). This will happen when v_0 or d_0 is not zero. This will be the case when the object of interest is already in motion, or the motion begins at some point other than at the origin of the coordinate system.

Virtual Physics

The Moving Man (Part 2)

Look at the Moving Man simulation again and this time use the *Charts* view. Again, vary the velocity and acceleration by sliding the red and green markers along the scales. Keeping the velocity marker near zero will make the effect of acceleration more obvious. Observe how the graphs of position, velocity, and acceleration vary with time. Note which are linear plots and which are not.

Click to view content (https://archive.cnx.org/specials/e2ca52af-8c6b-450e-ac2f-9300b38e8739/moving-man/)

GRASP CHECK

On a velocity versus time plot, what does the slope represent?

a. Acceleration

b. Displacement
c. Distance covered
d. Instantaneous velocity

GRASP CHECK

On a position versus time plot, what does the slope represent?

a. Acceleration
b. Displacement
c. Distance covered
d. Instantaneous velocity

The kinematic equations are applicable when you have constant acceleration.

1. $d = d_0 + \overline{v}t$, or $\overline{v} = \frac{d}{t}$ when $d_0 = 0$
2. $\overline{v} = \frac{v_0 + v_f}{2}$
3. $v = v_0 + at$, or $a = \frac{v}{t}$ when $v_0 = 0$
4. $d = d_0 + v_0 t + \frac{1}{2}at^2$, or $a = \frac{2d}{t^2}$ when $d_0 = 0$ and $v_0 = 0$
5. $v^2 = v_0^2 + 2a(d - d_0)$, or $a = \frac{2d}{t^2}$ when $d_0 = 0$ and $v_0 = 0$

Applying Kinematic Equations to Situations of Constant Acceleration

Problem-solving skills are essential to success in a science and life in general. The ability to apply broad physical principles, which are often represented by equations, to specific situations is a very powerful form of knowledge. It is much more powerful than memorizing a list of facts. Analytical skills and problem-solving abilities can be applied to new situations, whereas a list of facts cannot be made long enough to contain every possible circumstance. Essential analytical skills will be developed by solving problems in this text and will be useful for understanding physics and science in general throughout your life.

Problem-Solving Steps

While no single step-by-step method works for every problem, the following general procedures facilitate problem solving and make the answers more meaningful. A certain amount of creativity and insight are required as well.

1. *Examine the situation to determine which physical principles are involved. It is vital to draw a simple sketch at the outset.* Decide which direction is positive and note that on your sketch.
2. *Identify the knowns: Make a list of what information is given or can be inferred from the problem statement.* Remember, not all given information will be in the form of a number with units in the problem. If something starts *from rest*, we know the initial velocity is zero. If something *stops*, we know the final velocity is zero.
3. *Identify the unknowns: Decide exactly what needs to be determined in the problem.*
4. *Find an equation or set of equations that can help you solve the problem.* Your list of knowns and unknowns can help here. For example, if time is not needed or not given, then the fifth kinematic equation, which does not include time, could be useful.
5. *Insert the knowns along with their units into the appropriate equation and obtain numerical solutions complete with units.* This step produces the numerical answer; it also provides a check on units that can help you find errors. If the units of the answer are incorrect, then an error has been made.
6. *Check the answer to see if it is reasonable: Does it make sense?* This final step is extremely important because the goal of physics is to accurately describe nature. To see if the answer is reasonable, check its magnitude, its sign, and its units. Are the significant figures correct?

Summary of Problem Solving

- Determine the knowns and unknowns.
- Find an equation that expresses the unknown in terms of the knowns. More than one unknown means more than one equation is needed.
- Solve the equation or equations.

- Be sure units and significant figures are correct.
- Check whether the answer is reasonable.

FUN IN PHYSICS

Drag Racing

Figure 3.10 Smoke rises from the tires of a dragster at the beginning of a drag race. (Lt. Col. William Thurmond. Photo courtesy of U.S. Army.)

The object of the sport of drag racing is acceleration. Period! The races take place from a standing start on a straight one-quarter-mile (402 m) track. Usually two cars race side by side, and the winner is the driver who gets the car past the quarter-mile point first. At the finish line, the cars may be going more than 300 miles per hour (134 m/s). The driver then deploys a parachute to bring the car to a stop because it is unsafe to brake at such high speeds. The cars, called dragsters, are capable of accelerating at 26 m/s^2. By comparison, a typical sports car that is available to the general public can accelerate at about 5 m/s^2.

Several measurements are taken during each drag race:

- Reaction time is the time between the starting signal and when the front of the car crosses the starting line.
- Elapsed time is the time from when the vehicle crosses the starting line to when it crosses the finish line. The record is a little over 3 s.
- Speed is the average speed during the last 20 m before the finish line. The record is a little under 400 mph.

The video shows a race between two dragsters powered by jet engines. The actual race lasts about four seconds and is near the end of the video (https://openstax.org/l/28dragsters) .

GRASP CHECK

A dragster crosses the finish line with a velocity of 140 m/s. Assuming the vehicle maintained a constant acceleration from start to finish, what was its average velocity for the race?

a. 0 m/s
b. 35 m/s
c. 70 m/s
d. 140 m/s

WORKED EXAMPLE

Acceleration of a Dragster

The time it takes for a dragster to cross the finish line is unknown. The dragster accelerates from rest at 26 m/s^2 for a quarter mile (0.250 mi). What is the final velocity of the dragster?

Strategy

The equation $v^2 = v_0^2 + 2a(d - d_0)$ is ideally suited to this task because it gives the velocity from acceleration and displacement, without involving the time.

Solution

1. Convert miles to meters.

$$(0.250\ \text{mi}) \times \frac{1609\ \text{m}}{1\ \text{mi}} = 402\ \text{m} \tag{3.9}$$

2. Identify the known values. We know that $v_0 = 0$ since the dragster starts from rest, and we know that the distance traveled, $d - d_0$ is 402 m. Finally, the acceleration is constant at $a = 26.0\ \text{m/s}^2$.
3. Insert the knowns into the equation $v^2 = v_0^2 + 2a(d - d_0)$ and solve for v.

$$v^2 = 0 + 2\left(26.0\frac{\text{m}}{\text{s}^2}\right)(402\ \text{m}) = 2.09 \times 10^4\ \frac{\text{m}^2}{\text{s}^2} \tag{3.10}$$

Taking the square root gives us $v = \sqrt{2.09 \times 10^4\ \frac{\text{m}^2}{\text{s}^2}} = 145\ \frac{\text{m}}{\text{s}}$.

Discussion

145 m/s is about 522 km/hour or about 324 mi/h, but even this breakneck speed is short of the record for the quarter mile. Also, note that a square root has two values. We took the positive value because we know that the velocity must be in the same direction as the acceleration for the answer to make physical sense.

An examination of the equation $v^2 = v_0^2 + 2a(d - d_0)$ can produce further insights into the general relationships among physical quantities:

- The final velocity depends on the magnitude of the acceleration and the distance over which it applies.
- For a given acceleration, a car that is going twice as fast does not stop in twice the distance—it goes much further before it stops. This is why, for example, we have reduced speed zones near schools.

Practice Problems

6. Dragsters can reach a top speed of 145 m/s in only 4.45 s. Calculate the average acceleration for such a dragster.
   a. $-32.6\ \text{m/s}^2$
   b. $0\ \text{m/s}^2$
   c. $32.6\ \text{m/s}^2$
   d. $145\ \text{m/s}^2$

7. An Olympic-class sprinter starts a race with an acceleration of $4.50\ \text{m/s}^2$. Assuming she can maintain that acceleration, what is her speed 2.40 s later?
   a. 4.50 m/s
   b. 10.8 m/s
   c. 19.6 m/s
   d. 44.1 m/s

Constant Acceleration

In many cases, acceleration is not uniform because the force acting on the accelerating object is not constant over time. A situation that gives constant acceleration is the acceleration of falling objects. When air resistance is not a factor, all objects near Earth's surface fall with an acceleration of about $9.80\ \text{m/s}^2$. Although this value decreases slightly with increasing altitude, it may be assumed to be essentially constant. The value of $9.80\ \text{m/s}^2$ is labeled g and is referred to as **acceleration due to gravity**. Gravity is the force that causes nonsupported objects to accelerate downward—or, more precisely, toward the center of Earth. The magnitude of this force is called the weight of the object and is given by mg where m is the mass of the object (in kg). In places other than on Earth, such as the Moon or on other planets, g is not $9.80\ \text{m/s}^2$, but takes on other values. When using g for the acceleration a in a kinematic equation, it is usually given a negative sign because the acceleration due to gravity is downward.

WORK IN PHYSICS

Effects of Rapid Acceleration

Figure 3.11 Astronauts train using G Force Simulators. (NASA)

When in a vehicle that accelerates rapidly, you experience a force on your entire body that accelerates your body. You feel this force in automobiles and slightly more on amusement park rides. For example, when you ride in a car that turns, the car applies a force on your body to make you accelerate in the direction in which the car is turning. If enough force is applied, you will accelerate at 9.80 m/s^2. This is the same as the acceleration due to gravity, so this force is called one G.

One G is the force required to accelerate an object at the acceleration due to gravity at Earth's surface. Thus, one G for a paper cup is much less than one G for an elephant, because the elephant is much more massive and requires a greater force to make it accelerate at 9.80 m/s^2. For a person, a G of about 4 is so strong that his or her face will distort as the bones accelerate forward through the loose flesh. Other symptoms at extremely high Gs include changes in vision, loss of consciousness, and even death. The space shuttle produces about 3 Gs during takeoff and reentry. Some roller coasters and dragsters produce forces of around 4 Gs for their occupants. A fighter jet can produce up to 12 Gs during a sharp turn.

Astronauts and fighter pilots must undergo G-force training in simulators. The video (https://www.youtube.com/watch?v=n-8QHOUWECU) shows the experience of several people undergoing this training.

People, such as astronauts, who work with G forces must also be trained to experience zero G—also called free fall or weightlessness—which can cause queasiness. NASA has an aircraft that allows it occupants to experience about 25 s of free fall. The aircraft is nicknamed the *Vomit Comet*.

GRASP CHECK

A common way to describe acceleration is to express it in multiples of g, Earth's gravitational acceleration. If a dragster accelerates at a rate of 39.2 m/s^2, how many g's does the driver experience?

a. 1.5 g
b. 4.0 g
c. 10.5 g
d. 24.5 g

WORKED EXAMPLE

Falling Objects

A person standing on the edge of a high cliff throws a rock straight up with an initial velocity v_0 of 13 m/s.
(a) Calculate the position and velocity of the rock at 1.00, 2.00, and 3.00 seconds after it is thrown. Ignore the effect of air resistance.

Strategy

Sketch the initial velocity and acceleration vectors and the axes.

Figure 3.12 Initial conditions for rock thrown straight up.

List the knowns: time $t = 1.00$ s, 2.00 s, and 3.00 s; initial velocity $v_0 = 13$ m/s; acceleration $a = g = -9.80$ m/s²; and position $y_0 = 0$ m

List the unknowns: y_1, y_2, and y_3; v_1, v_2, and v_3—where 1, 2, 3 refer to times 1.00 s, 2.00 s, and 3.00 s

Choose the equations.

$$d = d_0 + v_0 t + \frac{1}{2}at^2 \text{ becomes } y = y_0 + v_0 t - \frac{1}{2}gt^2 \tag{3.11}$$

$$v = v_0 + at \text{ becomes } v = v_0 + -gt \tag{3.12}$$

These equations describe the unknowns in terms of knowns only.

Solution

$$y_1 = 0 + (13.0 \text{ m/s})(1.00 \text{ s}) + \frac{(-9.80\text{m/s}^2)(1.00 \text{ s})^2}{2} = 8.10 \text{ m}$$

$$y_2 = 0 + (13.0 \text{ m/s})(2.00 \text{ s}) + \frac{(-9.80\text{m/s}^2)(2.00 \text{ s})^2}{2} = 6.40 \text{ m}$$

$$y_3 = 0 + (13.0 \text{ m/s})(3.00 \text{ s}) + \frac{(-9.80\text{m/s}^2)(3.00 \text{ s})^2}{2} = -5.10 \text{ m}$$

$$v_1 = 13.0 \text{ m/s} + \left(-9.80\text{m/s}^2\right)(1.00 \text{ s}) = 3.20 \text{ m/s}$$

$$v_2 = 13.0 \text{ m/s} + \left(-9.80\text{m/s}^2\right)(2.00 \text{ s}) = -6.60 \text{ m/s}$$

$$v_3 = 13.0 \text{ m/s} + \left(-9.80\text{m/s}^2\right)(3.00 \text{ s}) = -16.4 \text{ m/s}$$

Discussion

The first two positive values for y show that the rock is still above the edge of the cliff, and the third negative y value shows that it has passed the starting point and is below the cliff. Remember that we set *up* to be positive. Any position with a positive value is above the cliff, and any velocity with a positive value is an upward velocity. The first value for v is positive, so the rock is still on the way up. The second and third values for v are negative, so the rock is on its way down.

(b) Make graphs of position versus time, velocity versus time, and acceleration versus time. Use increments of 0.5 s in your graphs.

Strategy

Time is customarily plotted on the *x*-axis because it is the independent variable. Position versus time will not be linear, so calculate points for 0.50 s, 1.50 s, and 2.50 s. This will give a curve closer to the true, smooth shape.

Solution

The three graphs are shown in Figure 3.13.

Figure 3.13

Discussion

- *y* vs. *t* does *not* represent the two-dimensional parabolic path of a trajectory. The path of the rock is straight up and straight down. The slope of a line tangent to the curve at any point on the curve equals the velocity at that point—i.e., the instantaneous velocity.

- Note that the *v* vs. *t* line crosses the vertical axis at the initial velocity and crosses the horizontal axis at the time when the rock changes direction and begins to fall back to Earth. This plot is linear because acceleration is constant.
- The *a* vs. *t* plot also shows that acceleration is constant; that is, it does not change with time.

Practice Problems

8. A cliff diver pushes off horizontally from a cliff and lands in the ocean $2.00\ \text{s}$ later. How fast was he going when he entered the water?

a. $0\ \text{m/s}$
b. $19.0\ \text{m/s}$
c. $19.6\ \text{m/s}$
d. $20.0\ \text{m/s}$

9. A girl drops a pebble from a high cliff into a lake far below. She sees the splash of the pebble hitting the water $2.00\ \text{s}$ later. How fast was the pebble going when it hit the water?

a. $9.80\ \text{m/s}$
b. $10.0\ \text{m/s}$
c. $19.6\ \text{m/s}$
d. $20.0\ \text{m/s}$

Check Your Understanding

10. Identify the four variables found in the kinematic equations.

a. Displacement, Force, Mass, and Time
b. Acceleration, Displacement, Time, and Velocity
c. Final Velocity, Force, Initial Velocity, and Mass
d. Acceleration, Final Velocity, Force, and Initial Velocity

11. Which of the following steps is always required to solve a kinematics problem?

a. Find the force acting on the body.
b. Find the acceleration of a body.
c. Find the initial velocity of a body.
d. Find a suitable kinematic equation and then solve for the unknown quantity.

12. Which of the following provides a correct answer for a problem that can be solved using the kinematic equations?

a. A body starts from rest and accelerates at $4\ \text{m/s}^2$ for $2\ \text{s}$. The body's final velocity is $8\ \text{m/s}$.
b. A body starts from rest and accelerates at $4\ \text{m/s}^2$ for $2\ \text{s}$. The body's final velocity is $16\ \text{m/s}$.
c. A body with a mass of $2\ \text{kg}$ is acted upon by a force of $4\ \text{N}$. The acceleration of the body is $2\ \text{m/s}^2$.
d. A body with a mass of $2\ \text{kg}$ is acted upon by a force of $4\ \text{N}$. The acceleration of the body is $0.5\ \text{m/s}^2$.

KEY TERMS

acceleration due to gravity acceleration of an object that is subject only to the force of gravity; near Earth's surface this acceleration is 9.80 m/s^2

average acceleration change in velocity divided by the time interval over which it changed

constant acceleration acceleration that does not change with respect to time

instantaneous acceleration rate of change of velocity at a specific instant in time

kinematic equations the five equations that describe motion in terms of time, displacement, velocity, and acceleration

negative acceleration acceleration in the negative direction

SECTION SUMMARY

3.1 Acceleration

- Acceleration is the rate of change of velocity and may be negative or positive.
- Average acceleration is expressed in m/s^2 and, in one dimension, can be calculated using $\overline{a} = \frac{\Delta v}{\Delta t} = \frac{v_f - v_0}{t_f - t_o}$.

3.2 Representing Acceleration with Equations and Graphs

- The kinematic equations show how time, displacement, velocity, and acceleration are related for objects in motion.
- In general, kinematic problems can be solved by identifying the kinematic equation that expresses the unknown in terms of the knowns.
- Displacement, velocity, and acceleration may be displayed graphically versus time.

KEY EQUATIONS

3.1 Acceleration

Average acceleration	$\overline{a} = \frac{\Delta v}{\Delta t} = \frac{v_f - v_0}{t_f - t_o}$
Average velocity	$\overline{v} = \frac{v_0 + v_f}{2}$
Velocity	$v = v_0 + at$, or when $v_o = 0$
Displacement	$d = d_0 + v_0 t + \frac{1}{2}at^2$, or $a = \frac{2d}{t^2}$ when $d_o = 0$ and $v_o = 0$
Acceleration	$v^2 = v_0^2 + 2a(d - d_0)$, or $a = \frac{v^2}{2d}$ when $d_o = 0$ and $v_o = 0$

3.2 Representing Acceleration with Equations and Graphs

Average velocity	$d = d_0 + \overline{v}t$, or $\overline{v} = \frac{d}{t}$ when $d_o = 0$

CHAPTER REVIEW

Concept Items

3.1 Acceleration

1. How can you use the definition of acceleration to explain the units in which acceleration is measured?
   a. Acceleration is the rate of change of velocity. Therefore, its unit is m/s^2.
   b. Acceleration is the rate of change of displacement. Therefore, its unit is m/s.
   c. Acceleration is the rate of change of velocity. Therefore, its unit is m^2/s.
   d. Acceleration is the rate of change of displacement. Therefore, its unit is m^2/s.

2. What are the SI units of acceleration?
   a. m^2/s
   b. cm^2/s
   c. m/s^2
   d. cm/s^2

3. Which of the following statements explains why a racecar going around a curve is accelerating, even if the speed is constant?
   a. The car is accelerating because the magnitude as well as the direction of velocity is changing.
   b. The car is accelerating because the magnitude of velocity is changing.
   c. The car is accelerating because the direction of velocity is changing.

d. The car is accelerating because neither the magnitude nor the direction of velocity is changing.

3.2 Representing Acceleration with Equations and Graphs

4. A student calculated the final velocity of a train that decelerated from 30.5 m/s and got an answer of –43.34 m/s. Which of the following might indicate that he made a mistake in his calculation?
   a. The sign of the final velocity is wrong.
   b. The magnitude of the answer is too small.
   c. There are too few significant digits in the answer.
   d. The units in the initial velocity are incorrect.

5. Create your own kinematics problem. Then, create a flow chart showing the steps someone would need to take to solve the problem.
   a. Acceleration
   b. Distance
   c. Displacement
   d. Force

6. Which kinematic equation would you use to find the velocity of a skydiver 2.0 s after she jumps from a plane and before she opens her parachute? Assume the positive direction is downward.
   a. $v = v_0 + at$
   b. $v = v_0 - at$
   c. $v^2 = {v_0}^2 + at$
   d. $v^2 = {v_0}^2 - at$

Critical Thinking Items

3.1 Acceleration

7. Imagine that a car is traveling from your left to your right at a constant velocity. Which two actions could the driver take that may be represented as (a) a velocity vector and an acceleration vector both pointing to the right and then (b) changing so the velocity vector points to the right and the acceleration vector points to the left?
   a. (a) Push down on the accelerator and then (b) push down again on the accelerator a second time.
   b. (a) Push down on the accelerator and then (b) push down on the brakes.
   c. (a) Push down on the brakes and then (b) push down on the brakes a second time.
   d. (a) Push down on the brakes and then (b) push down on the accelerator.

8. A motorcycle moving at a constant velocity suddenly accelerates at a rate of 4.0 m/s^2 to a speed of 35 m/s in 5.0 s. What was the initial speed of the motorcycle?
   a. -34 m/s
   b. -15 m/s
   c. 15 m/s
   d. 34 m/s

3.2 Representing Acceleration with Equations and Graphs

9. A student is asked to solve a problem:
   An object falls from a height for 2.0 s, at which point it is still 60 m above the ground. What will be the velocity of the object when it hits the ground?
   Which of the following provides the correct order of kinematic equations that can be used to solve the problem?
   a. First use $v^2 = {v_0}^2 + 2a(d - d_0)$, then use $v = v_0 + at$.
   b. First use $v = v_0 + at$, then use $v^2 = {v_0}^2 + 2a(d - d_0)$.
   c. First use $d = d_0 + v_0 t + \frac{1}{2}at^2$, then use $v = v_0 + at$.
   d. First use $v = v_0 + at$, then use $d - d_0 = v_0 t + \frac{1}{2}at^2$.

10. Skydivers are affected by acceleration due to gravity and by air resistance. Eventually, they reach a speed where the force of gravity is almost equal to the force of air resistance. As they approach that point, their acceleration decreases in magnitude to near zero.
    Part A. Describe the shape of the graph of the magnitude of the acceleration versus time for a falling skydiver.
    Part B. Describe the shape of the graph of the magnitude of the velocity versus time for a falling skydiver.
    Part C. Describe the shape of the graph of the magnitude of the displacement versus time for a falling skydiver.
    a. Part A. Begins with a nonzero y-intercept with a downward slope that levels off at zero; Part B. Begins at zero with an upward slope that decreases in magnitude until the curve levels off; Part C. Begins at zero with an upward slope that increases in magnitude until it becomes a positive constant
    b. Part A. Begins with a nonzero y-intercept with an upward slope that levels off at zero; Part B. Begins at zero with an upward slope that decreases in magnitude until the curve levels off; Part C. Begins at zero with an upward slope that increases in magnitude until it becomes a positive constant
    c. Part A. Begins with a nonzero y-intercept with a downward slope that levels off at zero; Part B. Begins at zero with a downward slope that

decreases in magnitude until the curve levels off; Part C. Begins at zero with an upward slope that increases in magnitude until it becomes a positive constant

d. Part A. Begins with a nonzero y-intercept with an upward slope that levels off at zero; Part B. Begins at zero with a downward slope that decreases in magnitude until the curve levels off; Part C. Begins at zero with an upward slope that increases in magnitude until it becomes a positive constant

11. Which graph in the previous problem has a positive slope?

a. Displacement versus time only
b. Acceleration versus time and velocity versus time
c. Velocity versus time and displacement versus time
d. Acceleration versus time and displacement versus time

Problems

3.1 Acceleration

12. The driver of a sports car traveling at 10.0 m/s steps down hard on the accelerator for 5.0 s and the velocity increases to 30.0 m/s. What was the average acceleration of the car during the 5.0 s time interval?

a. -1.0×102 m/s^2
b. -4.0 m/s^2
c. 4.0 m/s^2
d. 1.0×102 m/s^2

13. A girl rolls a basketball across a basketball court. The ball slowly decelerates at a rate of -0.20 m/s^2. If the initial velocity was 2.0 m/s and the ball rolled to a stop at 5.0 sec after 12:00 p.m., at what time did she start the ball rolling?

a. 0.1 seconds before noon
b. 0.1 seconds after noon
c. 5 seconds before noon
d. 5 seconds after noon

3.2 Representing Acceleration with Equations and Graphs

14. A swan on a lake gets airborne by flapping its wings and running on top of the water. If the swan must reach a velocity of 6.00 m/s to take off and it accelerates from rest at an average rate of 0.350 m/s^2, how far will it travel before becoming airborne?

a. -8.60 m
b. 8.60 m
c. -51.4 m
d. 51.4 m

15. A swimmer bounces straight up from a diving board and falls feet first into a pool. She starts with a velocity of 4.00 m/s and her takeoff point is 8 m above the pool. How long are her feet in the air?

a. 0.408 s
b. 0.816 s
c. 1.34 s
d. 1.75 s
e. 1.28 s

Performance Task

3.2 Representing Acceleration with Equations and Graphs

16. Design an experiment to measure displacement and elapsed time. Use the data to calculate final velocity, average velocity, acceleration, and acceleration.

Materials

- a small marble or ball bearing
- a garden hose
- a measuring tape
- a stopwatch or stopwatch software download
- a sloping driveway or lawn as long as the garden hose with a level area beyond

(a) How would you use the garden hose, stopwatch, marble, measuring tape, and slope to measure displacement and elapsed time? Hint—The marble is the accelerating object, and the length of the hose is total displacement.
(b) How would you use the displacement and time data to calculate velocity, average velocity, and acceleration? Which kinematic equations would you use?
(c) How would you use the materials, the measured and calculated data, and the flat area below the slope to determine the negative acceleration? What would you measure, and which kinematic equation would you use?

TEST PREP

Multiple Choice

3.1 Acceleration

17. Which variable represents displacement?
- a. a
- b. d
- c. t
- d. v

18. If a velocity increases from 0 to 20 m/s in 10 s, what is the average acceleration?
- a. 0.5 m/s^2
- b. 2 m/s^2
- c. 10 m/s^2
- d. 30 m/s^2

3.2 Representing Acceleration with Equations and Graphs

19. For the motion of a falling object, which graphs are straight lines?
- a. Acceleration versus time only
- b. Displacement versus time only
- c. Displacement versus time and acceleration versus time
- d. Velocity versus time and acceleration versus time

20. A bullet in a gun is accelerated from the firing chamber to the end of the barrel at an average rate of 6.30×10^5 m/s^2 for 8.10×10^{-4} s. What is the bullet's final velocity when it leaves the barrel, commonly known as the muzzle velocity?
- a. 7.79 m/s
- b. 51.0 m/s
- c. 510 m/s
- d. 1020 m/s

Short Answer

3.1 Acceleration

21. True or False—The vector for a negative acceleration points in the opposite direction when compared to the vector for a positive acceleration.
- a. True
- b. False

22. If a car decelerates from 20 m/s to 15 m/s in 5 s, what is Δv?
- a. -5 m/s
- b. -1 m/s
- c. 1 m/s
- d. 5 m/s

23. How is the vector arrow representing an acceleration of magnitude 3 m/s^2 different from the vector arrow representing a negative acceleration of magnitude 3 m/s^2?
- a. They point in the same direction.
- b. They are perpendicular, forming a 90° angle between each other.
- c. They point in opposite directions.
- d. They are perpendicular, forming a 270° angle between each other.

24. How long does it take to accelerate from 8.0 m/s to 20.0 m/s at a rate of acceleration of 3.0 m/s^2?
- a. 0.25 s
- b. 4.0 s
- c. 9.33 s
- d. 36 s

3.2 Representing Acceleration with Equations and Graphs

25. If a plot of displacement versus time is linear, what can be said about the acceleration?
- a. Acceleration is 0.
- b. Acceleration is a non-zero constant.
- c. Acceleration is positive.
- d. Acceleration is negative.

26.

True or False: —The image shows a velocity vs. time graph for a jet car. If you take the slope at any point on the graph, the jet car's acceleration will be 5.0 m/s^2.
- a. True
- b. False

27. When plotted on the blank plots, which answer choice would show the motion of an object that has uniformly accelerated

from 2 m/s to 8 m/s in 3 s?

a. The plot on the left shows a line from (0,2) to (3,8) while the plot on the right shows a line from (0,2) to (3,2).
b. The plot on the left shows a line from (0,2) to (3,8) while the plot on the right shows a line from (0,3) to (3,3).
c. The plot on the left shows a line from (0,8) to (3,2) while the plot on the right shows a line from (0,2) to (3,2).
d. The plot on the left shows a line from (0,8) to (3,2) while the plot on the right shows a line from (0,3) to (3,3).

28. When is a plot of velocity versus time a straight line and when is it a curved line?
a. It is a straight line when acceleration is changing and is a curved line when acceleration is constant.
b. It is a straight line when acceleration is constant and is a curved line when acceleration is changing.
c. It is a straight line when velocity is constant and is a curved line when velocity is changing.
d. It is a straight line when velocity is changing and is a curved line when velocity is constant.

Extended Response

3.1 Acceleration

29. A test car carrying a crash test dummy accelerates from 0 to 30 m/s and then crashes into a brick wall. Describe the direction of the initial acceleration vector and compare the initial acceleration vector's magnitude with respect to the acceleration magnitude at the moment of the crash.
a. The direction of the initial acceleration vector will point towards the wall, and its magnitude will be less than the acceleration vector of the crash.
b. The direction of the initial acceleration vector will point away from the wall, and its magnitude will be less than the vector of the crash.
c. The direction of the initial acceleration vector will point towards the wall, and its magnitude will be more than the acceleration vector of the crash.
d. The direction of the initial acceleration vector will point away from the wall, and its magnitude will be more than the acceleration vector of the crash.

30. A car accelerates from rest at a stop sign at a rate of 3.0 m/s^2 to a speed of 21.0 m/s, and then immediately begins to decelerate to a stop at the next stop sign at a rate of 4.0 m/s^2. How long did it take the car to travel from the first stop sign to the second stop sign? Show your work.
a. 1.7 seconds
b. 5.3 seconds
c. 7.0 seconds
d. 12 seconds

3.2 Representing Acceleration with Equations and Graphs

31. True or False: Consider an object moving with constant acceleration. The plot of displacement versus time for such motion is a curved line while the plot of displacement versus time squared is a straight line.
a. True
b. False

32. You throw a ball straight up with an initial velocity of 15.0 m/s. It passes a tree branch on the way up at a height of 7.00 m. How much additional time will pass before the ball passes the tree branch on the way back down?
a. 0.574 s
b. 0.956 s
c. 1.53 s
d. 1.91 s

CHAPTER 4
Forces and Newton's Laws of Motion

Figure 4.1 Newton's laws of motion describe the motion of the dolphin's path. (Credit: Jin Jang)

Chapter Outline

4.1 Force

4.2 Newton's First Law of Motion: Inertia

4.3 Newton's Second Law of Motion

4.4 Newton's Third Law of Motion

INTRODUCTION Isaac Newton (1642–1727) was a natural philosopher; a great thinker who combined science and philosophy to try to explain the workings of nature on Earth and in the universe. His laws of motion were just one part of the monumental work that has made him legendary. The development of Newton's laws marks the transition from the Renaissance period of history to the modern era. This transition was characterized by a revolutionary change in the way people thought about the physical universe. Drawing upon earlier work by scientists Galileo Galilei and Johannes Kepler, Newton's laws of motion allowed motion on Earth and in space to be predicted mathematically. In this chapter you will learn about force as well as Newton's first, second, and third laws of motion.

4.1 Force

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Differentiate between force, net force, and dynamics
- Draw a free-body diagram

Section Key Terms

dynamics	external force	force
free-body diagram	net external force	net force

Defining Force and Dynamics

Force is the cause of motion, and motion draws our attention. Motion itself can be beautiful, such as a dolphin jumping out of the water, the flight of a bird, or the orbit of a satellite. The study of motion is called kinematics, but kinematics describes only the way objects move—their velocity and their acceleration. **Dynamics** considers the forces that affect the motion of moving objects and systems. Newton's laws of motion are the foundation of dynamics. These laws describe the way objects speed up, slow down, stay in motion, and interact with other objects. They are also universal laws: they apply everywhere on Earth as well as in space.

A force pushes or pulls an object. The object being moved by a force could be an inanimate object, a table, or an animate object, a person. The pushing or pulling may be done by a person, or even the gravitational pull of Earth. Forces have different magnitudes and directions; this means that some forces are stronger than others and can act in different directions. For example, a cannon exerts a strong force on the cannonball that is launched into the air. In contrast, a mosquito landing on your arm exerts only a small force on your arm.

When multiple forces act on an object, the forces combine. Adding together all of the forces acting on an object gives the total force, or **net force**. An **external force** is a force that acts on an object within the system *from outside* the system. This type of force is different than an internal force, which acts between two objects that are both within the system. **The net external force** combines these two definitions; it is the total combined external force. We discuss further details about net force, external force, and net external force in the coming sections.

In mathematical terms, two forces acting in opposite directions have opposite *signs* (positive or negative). By convention, the negative sign is assigned to any movement to the left or downward. If two forces pushing in opposite directions are added together, the larger force will be somewhat canceled out by the smaller force pushing in the opposite direction. It is important to be consistent with your chosen coordinate system within a problem; for example, if negative values are assigned to the downward direction for velocity, then distance, force, and acceleration should also be designated as being negative in the downward direction.

Free-Body Diagrams and Examples of Forces

For our first example of force, consider an object hanging from a rope. This example gives us the opportunity to introduce a useful tool known as a **free-body diagram**. A free-body diagram represents the object being acted upon—that is, the free body—as a single point. Only the forces acting *on* the body (that is, external forces) are shown and are represented by vectors (which are drawn as arrows). These forces are the only ones shown because only external forces acting on the body affect its motion. We can ignore any internal forces within the body because they cancel each other out, as explained in the section on Newton's third law of motion. Free-body diagrams are very useful for analyzing forces acting on an object.

Figure 4.2 An object of mass, *m*, is held up by the force of tension.

Figure 4.2 shows the force of tension in the rope acting in the upward direction, opposite the force of gravity. The forces are indicated in the free-body diagram by an arrow pointing up, representing tension, and another arrow pointing down, representing gravity. In a free-body diagram, the lengths of the arrows show the relative magnitude (or strength) of the forces. Because forces are vectors, they add just like other vectors. Notice that the two arrows have equal lengths in Figure 4.2, which means that the forces of tension and weight are of equal magnitude. Because these forces of equal magnitude act in opposite directions, they are perfectly balanced, so they add together to give a net force of zero.

Not all forces are as noticeable as when you push or pull on an object. Some forces act without physical contact, such as the pull of a magnet (in the case of magnetic force) or the gravitational pull of Earth (in the case of gravitational force).

In the next three sections discussing Newton's laws of motion, we will learn about three specific types of forces: friction, the normal force, and the gravitational force. To analyze situations involving forces, we will create free-body diagrams to organize the framework of the mathematics for each individual situation.

TIPS FOR SUCCESS

Correctly drawing and labeling a free-body diagram is an important first step for solving a problem. It will help you visualize the problem and correctly apply the mathematics to solve the problem.

Check Your Understanding

1. What is kinematics?
   a. Kinematics is the study of motion.
   b. Kinematics is the study of the cause of motion.
   c. Kinematics is the study of dimensions.
   d. Kinematics is the study of atomic structures.

2. Do two bodies have to be in physical contact to exert a force upon one another?

a. No, the gravitational force is a field force and does not require physical contact to exert a force.
b. No, the gravitational force is a contact force and does not require physical contact to exert a force.
c. Yes, the gravitational force is a field force and requires physical contact to exert a force.
d. Yes, the gravitational force is a contact force and requires physical contact to exert a force.

3. What kind of physical quantity is force?
   a. Force is a scalar quantity.
   b. Force is a vector quantity.
   c. Force is both a vector quantity and a scalar quantity.
   d. Force is neither a vector nor a scalar quantity.

4. Which forces can be represented in a free-body diagram?
   a. Internal forces
   b. External forces
   c. Both internal and external forces
   d. A body that is not influenced by any force

4.2 Newton's First Law of Motion: Inertia

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Describe Newton's first law and friction, and
- Discuss the relationship between mass and inertia.

Section Key Terms

friction	inertia	law of inertia
mass	Newton's first law of motion	system

Newton's First Law and Friction

Newton's first law of motion states the following:

1. A body at rest tends to remain at rest.
2. A body in motion tends to remain in motion at a constant velocity unless acted on by a net external force. (Recall that *constant velocity* means that the body moves in a straight line and at a constant speed.)

At first glance, this law may seem to contradict your everyday experience. You have probably noticed that a moving object will usually slow down and stop unless some effort is made to keep it moving. The key to understanding why, for example, a sliding box slows down (seemingly on its own) is to first understand that a net external force acts on the box to make the box slow down. Without this net external force, the box would continue to slide at a constant velocity (as stated in Newton's first law of motion). What force acts on the box to slow it down? This force is called **friction**. Friction is an external force that acts opposite to the direction of motion (see Figure 4.3). Think of friction as a resistance to motion that slows things down.

Consider an air hockey table. When the air is turned off, the puck slides only a short distance before friction slows it to a stop. However, when the air is turned on, it lifts the puck slightly, so the puck experiences very little friction as it moves over the surface. With friction almost eliminated, the puck glides along with very little change in speed. On a frictionless surface, the puck would experience no net external force (ignoring air resistance, which is also a form of friction). Additionally, if we know enough about friction, we can accurately predict how quickly objects will slow down.

Now let's think about another example. A man pushes a box across a floor at constant velocity by applying a force of +50 N. (The positive sign indicates that, by convention, the direction of motion is to the right.) What is the force of friction that opposes the motion? The force of friction must be –50 N. Why? According to Newton's first law of motion, any object moving at constant velocity has no net external force acting upon it, which means that the sum of the forces acting on the object must be zero. The mathematical way to say that no net external force acts on an object is $\mathbf{F}_{\text{net}} = 0$ or $\Sigma\mathbf{F} = 0$. So if the man applies +50 N of force, then the force of friction must be –50 N for the two forces to add up to zero (that is, for the two forces to *cancel* each

other). Whenever you encounter the phrase *at constant velocity*, Newton's first law tells you that the net external force is zero.

Figure 4.3 For a box sliding across a floor, friction acts in the direction opposite to the velocity.

The force of friction depends on two factors: the coefficient of friction and the normal force. For any two surfaces that are in contact with one another, the coefficient of friction is a constant that depends on the nature of the surfaces. The normal force is the force exerted by a surface that pushes on an object in response to gravity pulling the object down. In equation form, the force of friction is

$$\mathbf{f} = \mu \mathbf{N}, \tag{4.1}$$

where μ is the coefficient of friction and **N** is the normal force. (The coefficient of friction is discussed in more detail in another chapter, and the normal force is discussed in more detail in the section *Newton's Third Law of Motion*.)

Recall from the section on Force that a net external force acts from outside on the object of interest. A more precise definition is that it acts on the **system** of interest. A system is one or more objects that you choose to study. It is important to define the system at the beginning of a problem to figure out which forces are external and need to be considered, and which are internal and can be ignored.

For example, in Figure 4.4 (a), two children push a third child in a wagon at a constant velocity. The system of interest is the wagon plus the small child, as shown in part (b) of the figure. The two children behind the wagon exert external forces on this system ($\mathbf{F}_1$, $\mathbf{F}_2$). Friction *f* acting at the axles of the wheels and at the surface where the wheels touch the ground two other external forces acting on the system. Two more external forces act on the system: the weight **W** of the system pulling down and the normal force **N** of the ground pushing up. Notice that the wagon is not accelerating vertically, so Newton's first law tells us that the normal force balances the weight. Because the wagon is moving forward at a constant velocity, the force of friction must have the same strength as the sum of the forces applied by the two children.

Figure 4.4 (a) The wagon and rider form a *system* that is acted on by external forces. (b) The two children pushing the wagon and child provide two external forces. Friction acting at the wheel axles and on the surface of the tires where they touch the ground provide an external force that act against the direction of motion. The weight **W** and the normal force **N** from the ground are two more external forces acting on the system. All external forces are represented in the figure by arrows. All of the external forces acting on the system add together, but because the wagon moves at a constant velocity, all of the forces must add up to zero.

Mass and Inertia

Inertia is the tendency for an object at rest to remain at rest, or for a moving object to remain in motion in a straight line with constant speed. This key property of objects was first described by Galileo. Later, Newton incorporated the concept of inertia into his first law, which is often referred to as the **law of inertia**.

As we know from experience, some objects have more inertia than others. For example, changing the motion of a large truck is more difficult than changing the motion of a toy truck. In fact, the inertia of an object is proportional to the mass of the object. **Mass** is a measure of the amount of matter (or *stuff*) in an object. The quantity or amount of matter in an object is determined by the number and types of atoms the object contains. Unlike weight (which changes if the gravitational force changes), mass does not depend on gravity. The mass of an object is the same on Earth, in orbit, or on the surface of the moon. In practice, it is very difficult to count and identify all of the atoms and molecules in an object, so mass is usually not determined this way. Instead, the mass of an object is determined by comparing it with the standard kilogram. Mass is therefore expressed in kilograms.

TIPS FOR SUCCESS

In everyday language, people often use the terms *weight* and *mass* interchangeably—but this is not correct. Weight is actually a force. (We cover this topic in more detail in the section *Newton's Second Law of Motion*.)

WATCH PHYSICS

Newton's First Law of Motion

This video contrasts the way we thought about motion and force in the time before Galileo's concept of inertia and Newton's first law of motion with the way we understand force and motion now.

Click to view content (https://www.khanacademy.org/embed_video?v=5-ZFOhHQS68)

GRASP CHECK

Before we understood that objects have a tendency to maintain their velocity in a straight line unless acted upon by a net force, people thought that objects had a tendency to stop on their own. This happened because a specific force was not yet understood. What was that force?

a. Gravitational force
b. Electrostatic force
c. Nuclear force
d. Frictional force

Virtual Physics

Forces and Motion—Basics

In this simulation, you will first explore net force by placing blue people on the left side of a tug-of-war rope and red people on the right side of the rope (by clicking people and dragging them with your mouse). Experiment with changing the number and size of people on each side to see how it affects the outcome of the match and the net force. Hit the "Go!" button to start the match, and the "reset all" button to start over.

Next, click on the Friction tab. Try selecting different objects for the person to push. Slide the *applied force* button to the right to apply force to the right, and to the left to apply force to the left. The force will continue to be applied as long as you hold down the button. See the arrow representing friction change in magnitude and direction, depending on how much force you apply. Try increasing or decreasing the friction force to see how this change affects the motion.

Click to view content (https://phet.colorado.edu/sims/html/forces-and-motion-basics/latest/forces-and-motion-basics_en.html)

GRASP CHECK

Click on the tab for the *Acceleration Lab and check the Sum of Forces* option. Push the box to the right and then release. Notice which direction the sum of forces arrow points after the person stops pushing the box and lets it continue moving to the right on its own. At this point, in which direction is the net force, the sum of forces, pointing? Why?

a. The net force acts to the right because the applied external force acted to the right.
b. The net force acts to the left because the applied external force acted to the left.
c. The net force acts to the right because the frictional force acts to the right.
d. The net force acts to the left because the frictional force acts to the left.

Check Your Understanding

5. What does Newton's first law state?
   a. A body at rest tends to remain at rest and a body in motion tends to remain in motion at a constant acceleration unless acted on by a net external force.
   b. A body at rest tends to remain at rest and a body in motion tends to remain in motion at a constant velocity unless acted on by a net external force.
   c. The rate of change of momentum of a body is directly proportional to the external force applied to the body.
   d. The rate of change of momentum of a body is inversely proportional to the external force applied to the body.

6. According to Newton's first law, a body in motion tends to remain in motion at a constant velocity. However, when you slide an object across a surface, the object eventually slows down and stops. Why?
   a. The object experiences a frictional force exerted by the surface, which opposes its motion.
   b. The object experiences the gravitational force exerted by Earth, which opposes its motion
   c. The object experiences an internal force exerted by the body itself, which opposes its motion.
   d. The object experiences a pseudo-force from the body in motion, which opposes its motion.

7. What is inertia?
   a. Inertia is an object's tendency to maintain its mass.
   b. Inertia is an object's tendency to remain at rest.
   c. Inertia is an object's tendency to remain in motion
   d. Inertia is an object's tendency to remain at rest or, if moving, to remain in motion.

8. What is mass? What does it depend on?
   a. Mass is the weight of an object, and it depends on the gravitational force acting on the object.
   b. Mass is the weight of an object, and it depends on the number and types of atoms in the object.
   c. Mass is the quantity of matter contained in an object, and it depends on the gravitational force acting on the object.
   d. Mass is the quantity of matter contained in an object, and it depends on the number and types of atoms in the object.

4.3 Newton's Second Law of Motion

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Describe Newton's second law, both verbally and mathematically
- Use Newton's second law to solve problems

Section Key Terms

freefall	Newton's second law of motion	weight

Describing Newton's Second Law of Motion

Newton's first law considered bodies at rest or bodies in motion at a constant velocity. The other state of motion to consider is when an object is moving with a changing velocity, which means a change in the speed and/or the direction of motion. This type of motion is addressed by **Newton's second law of motion**, which states how force causes changes in motion. Newton's second law of motion is used to calculate what happens in situations involving forces and motion, and it shows the mathematical relationship between force, mass, and *acceleration*. Mathematically, the second law is most often written as

$$\mathbf{F}_{\text{net}} = m\mathbf{a} \text{ or } \Sigma\mathbf{F} = m\mathbf{a}, \tag{4.2}$$

where $\mathbf{F}_{\text{net}}$ (or $\sum\mathbf{F}$) is the net external force, m is the mass of the system, and $\mathbf{a}$ is the acceleration. Note that $\mathbf{F}_{\text{net}}$ and $\sum\mathbf{F}$ are the same because the net external force is the sum of all the external forces acting on the system.

First, what do we mean by *a change in motion*? A change in motion is simply a change in velocity: the speed of an object can become slower or faster, the direction in which the object is moving can change, or both of these variables may change. A change in velocity means, by definition, that an acceleration has occurred. Newton's first law says that only a nonzero net external force can cause a change in motion, so a net external force must cause an acceleration. Note that acceleration can refer to slowing down or to speeding up. Acceleration can also refer to a change in the direction of motion with no change in speed, because acceleration is the change in velocity divided by the time it takes for that change to occur, *and* velocity is defined by speed *and* direction.

From the equation $\mathrm{F}_{\text{net}} = m\mathrm{a}$, we see that force is directly proportional to both mass and acceleration, which makes sense. To accelerate two objects from rest to the same velocity, you would expect more force to be required to accelerate the more massive object. Likewise, for two objects of the same mass, applying a greater force to one would accelerate it to a greater velocity.

Now, let's rearrange Newton's second law to solve for acceleration. We get

$$\mathbf{a} = \frac{\mathbf{F}_{\text{net}}}{m} \text{ or } \mathbf{a} = \frac{\Sigma\mathbf{F}}{m}. \tag{4.3}$$

In this form, we can see that acceleration is directly proportional to force, which we write as

$$\mathbf{a} \propto \mathbf{F}_{\text{net}}, \tag{4.4}$$

where the symbol $\propto$ means *proportional to*.

This proportionality mathematically states what we just said in words: acceleration is directly proportional to the net external

force. When two variables are directly proportional to each other, then if one variable doubles, the other variable must double. Likewise, if one variable is reduced by half, the other variable must also be reduced by half. In general, when one variable is multiplied by a number, the other variable is also multiplied by the same number. It seems reasonable that the acceleration of a system should be directly proportional to and in the same direction as the net external force acting on the system. An object experiences greater acceleration when acted on by a greater force.

It is also clear from the equation $\mathbf{a} = \mathbf{F}_{\text{net}}/m$ that acceleration is inversely proportional to mass, which we write as

$$\mathbf{a} \propto \frac{1}{m}. \tag{4.5}$$

Inversely proportional means that if one variable is multiplied by a number, the other variable must be *divided* by the same number. Now, it also seems reasonable that acceleration should be inversely proportional to the mass of the system. In other words, the larger the mass (the inertia), the smaller the acceleration produced by a given force. This relationship is illustrated in Figure 4.5, which shows that a given net external force applied to a basketball produces a much greater acceleration than when applied to a car.

Figure 4.5 The same force exerted on systems of different masses produces different accelerations. (a) A boy pushes a basketball to make a pass. The effect of gravity on the ball is ignored. (b) The same boy pushing with identical force on a stalled car produces a far smaller acceleration (friction is negligible). Note that the free-body diagrams for the ball and for the car are identical, which allows us to compare the two situations.

Applying Newton's Second Law

Before putting Newton's second law into action, it is important to consider units. The equation $\mathbf{F}_{\text{net}} = m\mathbf{a}$ is used to define the units of force in terms of the three basic units of mass, length, and time (recall that acceleration has units of length divided by time squared). The SI unit of force is called the newton (abbreviated N) and is the force needed to accelerate a 1-kg system at the rate of 1 m/s². That is, because $\mathbf{F}_{\text{net}} = m\mathbf{a}$, we have

$$1\ \text{N} = 1\ \text{kg} \times 1\text{m/s}^2 = 1\frac{\text{kg} \cdot \text{m}}{\text{s}^2}. \tag{4.6}$$

One of the most important applications of Newton's second law is to calculate **weight** (also known as the gravitational force), which is usually represented mathematically as $\mathbf{W}$. When people talk about gravity, they don't always realize that it is an acceleration. When an object is dropped, it accelerates toward the center of Earth. Newton's second law states that the net external force acting on an object is responsible for the acceleration of the object. If air resistance is negligible, the net external force on a falling object is only the gravitational force (i.e., the weight of the object).

Weight can be represented by a vector because it has a direction. Down is defined as the direction in which gravity pulls, so weight is normally considered a downward force. By using Newton's second law, we can figure out the equation for weight.

Consider an object with mass m falling toward Earth. It experiences only the force of gravity (i.e., the gravitational force or weight), which is represented by W. Newton's second law states that $\mathbf{F}_{\text{net}} = m\mathbf{a}$. Because the only force acting on the object is the gravitational force, we have $\mathbf{F}_{\text{net}} = \mathbf{W}$. We know that the acceleration of an object due to gravity is $\mathbf{g}$, so we have $\mathbf{a} = \mathbf{g}$. Substituting these two expressions into Newton's second law gives

$$\mathbf{W} = m\mathbf{g}. \tag{4.7}$$

This is the equation for weight—the gravitational force on a mass m. On Earth, $\mathbf{g} = 9.80\ \text{m/s}^2$, so the weight (disregarding for now the direction of the weight) of a 1.0-kg object on Earth is

$$\mathbf{W} = m\mathbf{g} = (1.0\ \text{kg})(9.80\ \text{m/s}^2) = 9.8\ \ \text{N}. \tag{4.8}$$

Although most of the world uses newtons as the unit of force, in the United States the most familiar unit of force is the pound (lb), where 1 N = 0.225 lb.

Recall that although gravity acts downward, it can be assigned a positive or negative value, depending on what the positive direction is in your chosen coordinate system. Be sure to take this into consideration when solving problems with weight. When the downward direction is taken to be negative, as is often the case, acceleration due to gravity becomes $\mathbf{g} = -9.8\ \text{m/s}^2$.

When the net external force on an object is its weight, we say that it is in **freefall**. In this case, the only force acting on the object is the force of gravity. On the surface of Earth, when objects fall downward toward Earth, they are never truly in freefall because there is always some upward force due to air resistance that acts on the object (and there is also the buoyancy force of air, which is similar to the buoyancy force in water that keeps boats afloat).

Gravity varies slightly over the surface of Earth, so the weight of an object depends very slightly on its location on Earth. Weight varies dramatically away from Earth's surface. On the moon, for example, the acceleration due to gravity is only $1.67\ \text{m/s}^2$. Because weight depends on the force of gravity, a 1.0-kg mass weighs 9.8 N on Earth and only about 1.7 N on the moon.

It is important to remember that weight and mass are very different, although they are closely related. Mass is the quantity of matter (how much *stuff*) in an object and does not vary, but weight is the gravitational force on an object and is proportional to the force of gravity. It is easy to confuse the two, because our experience is confined to Earth, and the weight of an object is essentially the same no matter where you are on Earth. Adding to the confusion, the terms mass and weight are often used interchangeably in everyday language; for example, our medical records often show our weight in kilograms, but never in the correct unit of newtons.

Snap Lab

Mass and Weight

In this activity, you will use a scale to investigate mass and weight.

- 1 bathroom scale
- 1 table

1. What do bathroom scales measure?
2. When you stand on a bathroom scale, what happens to the scale? It depresses slightly. The scale contains springs that compress in proportion to your weight—similar to rubber bands expanding when pulled.
3. The springs provide a measure of your weight (provided you are not accelerating). This is a force in newtons (or pounds). In most countries, the measurement is now divided by 9.80 to give a reading in kilograms, which is a of mass. The scale detects weight but is calibrated to display mass.
4. If you went to the moon and stood on your scale, would it detect the same *mass* as it did on Earth?

GRASP CHECK

While standing on a bathroom scale, push down on a table next to you. What happens to the reading? Why?

a. The reading increases because part of your weight is applied to the table and the table exerts a matching force on you that acts in the direction of your weight.
b. The reading increases because part of your weight is applied to the table and the table exerts a matching force on you that acts in the direction opposite to your weight.
c. The reading decreases because part of your weight is applied to the table and the table exerts a matching force on you that acts in the direction of your weight.
d. The reading decreases because part of your weight is applied to the table and the table exerts a matching force on

you that acts in the direction opposite to your weight.

TIPS FOR SUCCESS

Only *net external force* impacts the acceleration of an object. If more than one force acts on an object and you calculate the acceleration by using only one of these forces, you will not get the correct acceleration for that object.

WATCH PHYSICS

Newton's Second Law of Motion

This video reviews Newton's second law of motion and how net external force and acceleration relate to one another and to mass. It also covers units of force, mass, and acceleration, and reviews a worked-out example.

Click to view content (https://www.khanacademy.org/embed_video?v=ou9YMWlJgkE)

GRASP CHECK

True or False—If you want to reduce the acceleration of an object to half its original value, then you would need to reduce the net external force by half.

a. True
b. False

WORKED EXAMPLE

What Acceleration Can a Person Produce when Pushing a Lawn Mower?

Suppose that the net external force (push minus friction) exerted on a lawn mower is 51 N parallel to the ground. The mass of the mower is 240 kg. What is its acceleration?

Figure 4.6

Strategy

Because $\mathbf{F}_{\text{net}}$ and m are given, the acceleration can be calculated directly from Newton's second law: $\mathbf{F}_{\text{net}} = m\mathbf{a}$.

Solution

Solving Newton's second law for the acceleration, we find that the magnitude of the acceleration, $\mathbf{a}$, is $\mathbf{a} = \frac{\mathbf{F}_{\text{net}}}{m}$. Entering the given values for net external force and mass gives

$$\mathbf{a} = \frac{51\ \text{N}}{240\ \text{kg}} \tag{4.9}$$

Inserting the units $\text{kg} \cdot \text{m/s}^2$ for N yields

$$\mathbf{a} = \frac{51\ \text{kg} \cdot \text{m/s}^2}{240\ \text{kg}} = 0.21\ \text{m/s}^2. \tag{4.10}$$

Discussion

The acceleration is in the same direction as the net external force, which is parallel to the ground and to the right. There is no information given in this example about the individual external forces acting on the system, but we can say something about their relative magnitudes. For example, the force exerted by the person pushing the mower must be greater than the friction opposing the motion, because we are given that the net external force is in the direction in which the person pushes. Also, the vertical forces must cancel if there is no acceleration in the vertical direction (the mower is moving only horizontally). The acceleration found is reasonable for a person pushing a mower; the mower's speed must increase by 0.21 m/s every second, which is possible. The time during which the mower accelerates would not be very long because the person's top speed would soon be reached. At this point, the person could push a little less hard, because he only has to overcome friction.

WORKED EXAMPLE

What Rocket Thrust Accelerates This Sled?

Prior to manned space flights, rocket sleds were used to test aircraft, missile equipment, and physiological effects on humans at high accelerations. Rocket sleds consisted of a platform mounted on one or two rails and propelled by several rockets. Calculate the magnitude of force exerted by each rocket, called its thrust, **T**, for the four-rocket propulsion system shown below. The sled's initial acceleration is $49\ \text{m/s}^2$, the mass of the system is 2,100 kg, and the force of friction opposing the motion is 650 N.

Figure 4.7

Strategy

The system of interest is the rocket sled. Although forces act vertically on the system, they must cancel because the system does not accelerate vertically. This leaves us with only horizontal forces to consider. We'll assign the direction to the right as the positive direction. See the free-body diagram in Figure 4.8.

Solution

We start with Newton's second law and look for ways to find the thrust **T** of the engines. Because all forces and acceleration are along a line, we need only consider the magnitudes of these quantities in the calculations. We begin with

$$\mathbf{F}_{\text{net}} = m\mathbf{a}, \tag{4.11}$$

where $\mathbf{F}_{\text{net}}$ is the net external force in the horizontal direction. We can see from Figure 4.8 that the engine thrusts are in the same direction (which we call the positive direction), whereas friction opposes the thrust. In equation form, the net external force is

$$\mathbf{F}_{\text{net}} = 4\mathbf{T} - \mathbf{f}. \tag{4.12}$$

Newton's second law tells us that $\mathbf{F}_{\text{net}} = m\mathbf{a}$, so we get

$$m\mathbf{a} = 4\mathbf{T} - \mathbf{f}. \tag{4.13}$$

After a little algebra, we solve for the total thrust $4\mathbf{T}$:

$$4\mathbf{T} = m\mathbf{a} + \mathbf{f}, \tag{4.14}$$

which means that the individual thrust is

$$\mathbf{T} = \frac{m\mathbf{a} + \mathbf{f}}{4}. \tag{4.15}$$

Inserting the known values yields

$$\mathbf{T} = \frac{(2100\text{ kg})(49\text{ m/s}^2) + 650\text{ N}}{4} = 2.6 \times 10^4\text{ N}. \tag{4.16}$$

Discussion

The numbers are quite large, so the result might surprise you. Experiments such as this were performed in the early 1960s to test the limits of human endurance and to test the apparatus designed to protect fighter pilots in emergency ejections. Speeds of 1000 km/h were obtained, with accelerations of 45 **g**. (Recall that g, the acceleration due to gravity, is 9.80 m/s^2. An acceleration of 45 **g** is $45 \times 9.80\text{ m/s}^2$, which is approximately 440 m/s^2.) Living subjects are no longer used, and land speeds of 10,000 km/h have now been obtained with rocket sleds. In this example, as in the preceding example, the system of interest is clear. We will see in later examples that choosing the system of interest is crucial—and that the choice is not always obvious.

Practice Problems

9. If 1 N is equal to 0.225 lb, how many pounds is 5 N of force?
- a. 0.045 lb
- b. 1.125 lb
- c. 2.025 lb
- d. 5.000 lb

10. How much force needs to be applied to a 5-kg object for it to accelerate at 20 m/s²?
- a. 1 N
- b. 10 N
- c. 100 N
- d. 1,000 N

Check Your Understanding

11. What is the mathematical statement for Newton's second law of motion?
- a. $F = ma$
- b. $F = 2ma$
- c. $F = \frac{m}{a}$
- d. $F = ma^2$

12. Newton's second law describes the relationship between which quantities?
- a. Force, mass, and time
- b. Force, mass, and displacement
- c. Force, mass, and velocity
- d. Force, mass, and acceleration

13. What is acceleration?
- a. Acceleration is the rate at which displacement changes.
- b. Acceleration is the rate at which force changes.
- c. Acceleration is the rate at which velocity changes.

d. Acceleration is the rate at which mass changes.

4.4 Newton's Third Law of Motion

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Describe Newton's third law, both verbally and mathematically
- Use Newton's third law to solve problems

Section Key Terms

Newton's third law of motion	normal force	tension	thrust

Describing Newton's Third Law of Motion

If you have ever stubbed your toe, you have noticed that although your toe initiates the impact, the surface that you stub it on exerts a force back on your toe. Although the first thought that crosses your mind is probably "ouch, that hurt" rather than "this is a great example of Newton's third law," both statements are true.

This is exactly what happens whenever one object exerts a force on another—each object experiences a force that is the same strength as the force acting on the other object but that acts in the opposite direction. Everyday experiences, such as stubbing a toe or throwing a ball, are all perfect examples of Newton's third law in action.

Newton's third law of motion states that whenever a first object exerts a force on a second object, the first object experiences a force equal in magnitude but opposite in direction to the force that it exerts.

Newton's third law of motion tells us that forces always occur in pairs, and one object cannot exert a force on another without experiencing the same strength force in return. We sometimes refer to these force pairs as *action-reaction* pairs, where the force exerted is the action, and the force experienced in return is the reaction (although which is which depends on your point of view).

Newton's third law is useful for figuring out which forces are external to a system. Recall that identifying external forces is important when setting up a problem, because the external forces must be added together to find the net force.

We can see Newton's third law at work by looking at how people move about. Consider a swimmer pushing off from the side of a pool, as illustrated in Figure 4.8. She pushes against the pool wall with her feet and accelerates in the direction opposite to her push. The wall has thus exerted on the swimmer a force of equal magnitude but in the direction opposite that of her push. You might think that two forces of equal magnitude but that act in opposite directions would cancel, *but they do not because they act on different systems.*

In this case, there are two different systems that we could choose to investigate: the swimmer or the wall. If we choose the swimmer to be the system of interest, as in the figure, then $F_{\text{wall on feet}}$ is an external force on the swimmer and affects her motion. Because acceleration is in the same direction as the net external force, the swimmer moves in the direction of $F_{\text{wall on feet}}$. Because the swimmer is our system (or object of interest) and not the wall, we do not need to consider the force $F_{\text{feet on wall}}$ because it originates *from* the swimmer rather than *acting on* the swimmer. Therefore, $F_{\text{feet on wall}}$ does not directly affect the motion of the system and does not cancel $F_{\text{wall on feet}}$. Note that the swimmer pushes in the direction opposite to the direction in which she wants to move.

Figure 4.8 When the swimmer exerts a force $\mathbf{F}_{\text{feet on wall}}$ on the wall, she accelerates in the direction opposite to that of her push. This means that the net external force on her is in the direction opposite to $\mathbf{F}_{\text{feet on wall}}$. This opposition is the result of Newton's third law of motion, which dictates that the wall exerts a force $\mathbf{F}_{\text{wall on feet}}$ on the swimmer that is equal in magnitude but that acts in the direction opposite to the force that the swimmer exerts on the wall.

Other examples of Newton's third law are easy to find. As a teacher paces in front of a whiteboard, he exerts a force backward on the floor. The floor exerts a reaction force in the forward direction on the teacher that causes him to accelerate forward. Similarly, a car accelerates because the ground pushes forward on the car's wheels in reaction to the car's wheels pushing backward on the ground. You can see evidence of the wheels pushing backward when tires spin on a gravel road and throw rocks backward.

Another example is the force of a baseball as it makes contact with the bat. Helicopters create lift by pushing air down, creating an upward reaction force. Birds fly by exerting force on air in the direction opposite that in which they wish to fly. For example, the wings of a bird force air downward and backward in order to get lift and move forward. An octopus propels itself forward in the water by ejecting water backward through a funnel in its body, which is similar to how a jet ski is propelled. In these examples, the octopus or jet ski push the water backward, and the water, in turn, pushes the octopus or jet ski forward.

Applying Newton's Third Law

Forces are classified and given names based on their source, how they are transmitted, or their effects. In previous sections, we discussed the forces called *push*, *weight*, and *friction*. In this section, applying Newton's third law of motion will allow us to explore three more forces: the **normal force**, **tension**, and **thrust**. However, because we haven't yet covered vectors in depth, we'll only consider one-dimensional situations in this chapter. Another chapter will consider forces acting in two dimensions.

The gravitational force (or weight) acts on objects at all times and everywhere on Earth. We know from Newton's second law that a net force produces an acceleration; so, why is everything not in a constant state of freefall toward the center of Earth? The answer is the normal force. The normal force is the force that a surface applies to an object to support the weight of that object; it acts perpendicular to the surface upon which the object rests. If an object on a flat surface is not accelerating, the net external force is zero, and the normal force has the same magnitude as the weight of the system but acts in the opposite direction. In equation form, we write that

$$\mathbf{N} = m\mathbf{g}. \tag{4.17}$$

Note that this equation is only true for a horizontal surface.

The word *tension* comes from the Latin word meaning *to stretch*. Tension is the force along the length of a flexible connector, such as a string, rope, chain, or cable. Regardless of the type of connector attached to the object of interest, one must remember that the connector can only pull (or *exert tension*) in the direction *parallel* to its length. Tension is a pull that acts parallel to the connector, and that acts in opposite directions at the two ends of the connector. This is possible because a flexible connector is simply a long series of action-reaction forces, except at the two ends where outside objects provide one member of the action-reaction forces.

Consider a person holding a mass on a rope, as shown in Figure 4.9.

Figure 4.9 When a perfectly flexible connector (one requiring no force to bend it) such as a rope transmits a force **T**, this force must be parallel to the length of the rope, as shown. The pull that such a flexible connector exerts is a tension. Note that the rope pulls with equal magnitude force but in opposite directions to the hand and to the mass (neglecting the weight of the rope). This is an example of Newton's third law. The rope is the medium that transmits forces of equal magnitude between the two objects but that act in opposite directions.

Tension in the rope must equal the weight of the supported mass, as we can prove by using Newton's second law. If the 5.00 kg mass in the figure is stationary, then its acceleration is zero, so $\mathbf{F}_{\text{net}} = 0$. The only external forces acting on the mass are its weight **W** and the tension **T** supplied by the rope. Summing the external forces to find the net force, we obtain

$$\mathbf{F}_{\text{net}} = \mathbf{T} - \mathbf{W} = 0, \tag{4.18}$$

where **T** and **W** are the magnitudes of the tension and weight, respectively, and their signs indicate direction, with up being positive. By substituting $m\mathbf{g}$ for $\mathbf{F}_{\text{net}}$ and rearranging the equation, the tension equals the weight of the supported mass, just as you would expect

$$\mathbf{T} = \mathbf{W} = m\mathbf{g}. \tag{4.19}$$

For a 5.00-kg mass (neglecting the mass of the rope), we see that

$$\mathbf{T} = m\mathbf{g} = (5.00\ \text{kg})(9.80\ \text{m/s}^2) = 49.0\ \text{N}. \tag{4.20}$$

Another example of Newton's third law in action is thrust. Rockets move forward by expelling gas backward at a high velocity. This means that the rocket exerts a large force backward on the gas in the rocket combustion chamber, and the gas, in turn, exerts a large force forward on the rocket in response. This reaction force is called *thrust*.

TIPS FOR SUCCESS

A common misconception is that rockets propel themselves by pushing on the ground or on the air behind them. They actually work better in a vacuum, where they can expel exhaust gases more easily.

LINKS TO PHYSICS

Math: Problem-Solving Strategy for Newton's Laws of Motion

The basics of problem solving, presented earlier in this text, are followed here with specific strategies for applying Newton's laws of motion. These techniques also reinforce concepts that are useful in many other areas of physics.

First, identify the physical principles involved. If the problem involves forces, then Newton's laws of motion are involved, and it

is important to draw a careful sketch of the situation. An example of a sketch is shown in Figure 4.10. Next, as in Figure 4.10, use vectors to represent all forces. Label the forces carefully, and make sure that their lengths are proportional to the magnitude of the forces and that the arrows point in the direction in which the forces act.

Figure 4.10 (a) A sketch of Tarzan hanging motionless from a vine. (b) Arrows are used to represent all forces. **T** is the tension exerted on Tarzan by the vine, $\mathbf{F}_\mathrm{T}$ is the force exerted on the vine by Tarzan, and **W** is Tarzan's weight (i.e., the force exerted on Tarzan by Earth's gravity). All other forces, such as a nudge of a breeze, are assumed to be negligible. (c) Suppose we are given Tarzan's mass and asked to find the tension in the vine. We define the system of interest as shown and draw a free-body diagram, as shown in (d). $\mathbf{F}_\mathrm{T}$ is no longer shown because it does not act on the system of interest; rather, $\mathbf{F}_\mathrm{T}$ acts on the outside world. (d) The free-body diagram shows only the external forces acting on Tarzan. For these to sum to zero, we must have $\mathbf{T} = \mathbf{W}$.

Next, make a list of knowns and unknowns and assign variable names to the quantities given in the problem. Figure out which variables need to be calculated; these are the unknowns. Now carefully define the system: which objects are of interest for the problem. This decision is important, because Newton's second law involves only external forces. Once the system is identified, it's possible to see which forces are external and which are internal (see Figure 4.10).

If the system acts on an object outside the system, then you know that the outside object exerts a force of equal magnitude but in the opposite direction on the system.

A diagram showing the system of interest and all the external forces acting on it is called a free-body diagram. Only external forces are shown on free-body diagrams, not acceleration or velocity. Figure 4.10 shows a free-body diagram for the system of interest.

After drawing a free-body diagram, apply Newton's second law to solve the problem. This is done in Figure 4.10 for the case of Tarzan hanging from a vine. When external forces are clearly identified in the free-body diagram, translate the forces into equation form and solve for the unknowns. Note that forces acting in opposite directions have opposite signs. By convention, forces acting downward or to the left are usually negative.

GRASP CHECK

If a problem has more than one system of interest, more than one free-body diagram is required to describe the external forces acting on the different systems.

a. True
b. False

WATCH PHYSICS

Newton's Third Law of Motion

This video explains Newton's third law of motion through examples involving push, normal force, and thrust (the force that propels a rocket or a jet).

Click to view content (https://www.openstax.org/l/astronaut)

GRASP CHECK

If the astronaut in the video wanted to move upward, in which direction should he throw the object? Why?

a. He should throw the object upward because according to Newton's third law, the object will then exert a force on him in the same direction (i.e., upward).
b. He should throw the object upward because according to Newton's third law, the object will then exert a force on him in the opposite direction (i.e., downward).
c. He should throw the object downward because according to Newton's third law, the object will then exert a force on him in the opposite direction (i.e., upward).
d. He should throw the object downward because according to Newton's third law, the object will then exert a force on him in the same direction (i.e., downward).

WORKED EXAMPLE

An Accelerating Subway Train

A physics teacher pushes a cart of demonstration equipment to a classroom, as in Figure 4.11. Her mass is 65.0 kg, the cart's mass is 12.0 kg, and the equipment's mass is 7.0 kg. To push the cart forward, the teacher's foot applies a force of 150 N in the opposite direction (backward) on the floor. Calculate the acceleration produced by the teacher. The force of friction, which opposes the motion, is 24.0 N.

Figure 4.11

Strategy

Because they accelerate together, we define the system to be the teacher, the cart, and the equipment. The teacher pushes backward with a force $\mathbf{F}_{foot}$ of 150 N. According to Newton's third law, the floor exerts a forward force $\mathbf{F}_{floor}$ of 150 N on the system. Because all motion is horizontal, we can assume that no net force acts in the vertical direction, and the problem becomes one dimensional. As noted in the figure, the friction f opposes the motion and therefore acts opposite the direction of $\mathbf{F}_{floor}$.

We should not include the forces $\mathbf{F}_{\text{teacher}}$, $\mathbf{F}_{\text{cart}}$, or $\mathbf{F}_{\text{foot}}$ because these are exerted *by* the system, not *on* the system. We find the net external force by adding together the external forces acting on the system (see the free-body diagram in the figure) and then use Newton's second law to find the acceleration.

Solution

Newton's second law is

$$\mathbf{a} = \frac{\mathbf{F}_{\text{net}}}{m}. \tag{4.21}$$

The net external force on the system is the sum of the external forces: the force of the floor acting on the teacher, cart, and equipment (in the horizontal direction) and the force of friction. Because friction acts in the opposite direction, we assign it a negative value. Thus, for the net force, we obtain

$$\mathbf{F}_{\text{net}} = \mathbf{F}_{\text{floor}} - \mathbf{f} = 150\text{ N} - 24.0\text{ N} = 126\text{ N}. \tag{4.22}$$

The mass of the system is the sum of the mass of the teacher, cart, and equipment.

$$m = (65.0 + 12.0 + 7.0)\text{ kg} = 84\text{ kg} \tag{4.23}$$

Insert these values of net F and m into Newton's second law to obtain the acceleration of the system.

$$\begin{aligned} \mathbf{a} &= \frac{\mathbf{F}_{\text{net}}}{m} \\ a &= \frac{126\text{ N}}{84\text{ kg}} = 1.5\text{ m/s}^2 \end{aligned} \tag{4.24}$$

$$F_1 < F_2 \tag{4.25}$$

Discussion

None of the forces between components of the system, such as between the teacher's hands and the cart, contribute to the net external force because they are internal to the system. Another way to look at this is to note that the forces between components of a system cancel because they are equal in magnitude and opposite in direction. For example, the force exerted by the teacher on the cart is of equal magnitude but in the opposite direction of the force exerted by the cart on the teacher. In this case, both forces act on the same system, so they cancel. Defining the system was crucial to solving this problem.

Practice Problems

14. What is the equation for the normal force for a body with mass m that is at rest on a horizontal surface?

a. N = m
b. N = mg
c. N = mv
d. N = g

15. An object with mass m is at rest on the floor. What is the magnitude and direction of the normal force acting on it?

a. N = mv in upward direction
b. N = mg in upward direction
c. N = mv in downward direction
d. N = mg in downward direction

Check Your Understanding

16. What is Newton's third law of motion?

a. Whenever a first body exerts a force on a second body, the first body experiences a force that is twice the magnitude and acts in the direction of the applied force.
b. Whenever a first body exerts a force on a second body, the first body experiences a force that is equal in magnitude and acts in the direction of the applied force.
c. Whenever a first body exerts a force on a second body, the first body experiences a force that is twice the magnitude but acts in the direction opposite the direction of the applied force.
d. Whenever a first body exerts a force on a second body, the first body experiences a force that is equal in magnitude but

acts in the direction opposite the direction of the applied force.

17. Considering Newton's third law, why don't two equal and opposite forces cancel out each other?
   a. Because the two forces act in the same direction
   b. Because the two forces have different magnitudes
   c. Because the two forces act on different systems
   d. Because the two forces act in perpendicular directions

KEY TERMS

dynamics the study of how forces affect the motion of objects and systems

external force a force acting on an object or system that originates outside of the object or system

force a push or pull on an object with a specific magnitude and direction; can be represented by vectors; can be expressed as a multiple of a standard force

free-body diagram a diagram showing all external forces acting on a body

freefall a situation in which the only force acting on an object is the force of gravity

friction an external force that acts in the direction opposite to the direction of motion

inertia the tendency of an object at rest to remain at rest, or for a moving object to remain in motion in a straight line and at a constant speed

law of inertia Newton's first law of motion: a body at rest remains at rest or, if in motion, remains in motion at a constant speed in a straight line, unless acted on by a net external force; also known as the law of inertia

mass the quantity of matter in a substance; measured in kilograms

net external force the sum of all external forces acting on an object or system

net force the sum of all forces acting on an object or system

Newton's first law of motion a body at rest remains at rest or, if in motion, remains in motion at a constant speed in a straight line, unless acted on by a net external force; also known as the law of inertia

Newton's second law of motion the net external force, $\mathbf{F}_{\text{net}}$, on an object is proportional to and in the same direction as the acceleration of the object, **a**, and also proportional to the object's mass, m; defined mathematically as $\mathbf{F}_{\text{net}} = m\mathbf{a}$ or $\Sigma\mathbf{F} = m\mathbf{a}$.

Newton's third law of motion when one body exerts a force on a second body, the first body experiences a force that is equal in magnitude and opposite in direction to the force that it exerts

normal force the force that a surface applies to an object; acts perpendicular and away from the surface with which the object is in contact

system one or more objects of interest for which only the forces acting on them from the outside are considered, but not the forces acting between them or inside them

tension a pulling force that acts along a connecting medium, especially a stretched flexible connector, such as a rope or cable; when a rope supports the weight of an object, the force exerted on the object by the rope is called tension

thrust a force that pushes an object forward in response to the backward ejection of mass by the object; rockets and airplanes are pushed forward by a thrust reaction force in response to ejecting gases backward

weight the force of gravity, **W**, acting on an object of mass m; defined mathematically as $\mathbf{W} = m\mathbf{g}$, where **g** is the magnitude and direction of the acceleration due to gravity

SECTION SUMMARY

4.1 Force

- Dynamics is the study of how forces affect the motion of objects and systems.
- Force is a push or pull that can be defined in terms of various standards. It is a vector and so has both magnitude and direction.
- External forces are any forces outside of a body that act on the body. A free-body diagram is a drawing of all external forces acting on a body.

4.2 Newton's First Law of Motion: Inertia

- Newton's first law states that a body at rest remains at rest or, if moving, remains in motion in a straight line at a constant speed, unless acted on by a net external force. This law is also known as the law of inertia.
- Inertia is the tendency of an object at rest to remain at rest or, if moving, to remain in motion at constant velocity. Inertia is related to an object's mass.
- Friction is a force that opposes motion and causes an object or system to slow down.
- Mass is the quantity of matter in a substance.

4.3 Newton's Second Law of Motion

- Acceleration is a change in velocity, meaning a change in speed, direction, or both.
- An external force acts on a system from outside the system, as opposed to internal forces, which act between components within the system.
- Newton's second law of motion states that the acceleration of a system is directly proportional to and in the same direction as the net external force acting on the system, and inversely proportional to the system's mass.
- In equation form, Newton's second law of motion is $\mathbf{F}_{\text{net}} = m\mathbf{a}$ or $\Sigma\mathbf{F} = m\mathbf{a}$. This is sometimes written as $\mathbf{a} = \frac{\mathbf{F}_{\text{net}}}{m}$ or $\mathbf{a} = \frac{\Sigma\mathbf{F}}{m}$.
- The weight of an object of mass m is the force of gravity that acts on it. From Newton's second law, weight is

given by $\mathbf{W} = m\mathbf{g}$.

- If the only force acting on an object is its weight, then the object is in freefall.

4.4 Newton's Third Law of Motion

- Newton's third law of motion states that when one body exerts a force on a second body, the first body experiences a force that is equal in magnitude and opposite in direction to the force that it exerts.
- When an object rests on a surface, the surface applies a force on the object that opposes the weight of the object. This force acts perpendicular to the surface and is called the normal force.
- The pulling force that acts along a stretched flexible connector, such as a rope or cable, is called tension. When a rope supports the weight of an object at rest, the tension in the rope is equal to the weight of the object.
- Thrust is a force that pushes an object forward in response to the backward ejection of mass by the object. Rockets and airplanes are pushed forward by thrust.

KEY EQUATIONS

4.2 Newton's First Law of Motion: Inertia

Newton's first law of motion	$\mathbf{F}_{\text{net}} = 0$ or $\Sigma\mathbf{F} = 0$

4.3 Newton's Second Law of Motion

Newton's second law of motion	$\mathbf{F}_{\text{net}} = m\mathbf{a}$ or $\Sigma\mathbf{F} = m\mathbf{a}$
Newton's second law of motion to solve acceleration	$\mathbf{a} = \frac{\mathbf{F}_{\text{net}}}{m}$ or $\mathbf{a} = \frac{\Sigma\mathbf{F}}{m}$
Newton's second law of motion to solve weight	$\mathbf{W} = m\mathbf{g}$

4.4 Newton's Third Law of Motion

normal force for a nonaccelerating horizontal surface	$\mathbf{N} = m\mathbf{g}$
tension for an object at rest	$\mathbf{T} = m\mathbf{g}$

CHAPTER REVIEW

Concept Items

4.1 Force

1. What is dynamics?
   a. Dynamics is the study of internal forces.
   b. Dynamics is the study of forces and their effect on motion.
   c. Dynamics describes the motion of points, bodies, and systems without consideration of the cause of motion.
   d. Dynamics describes the effect of forces on each other.

2. Two forces acting on an object are perpendicular to one another. How would you draw these in a free-body diagram?
   a. The two force arrows will be drawn at a right angle to one another.
   b. The two force arrows will be pointing in opposite directions.
   c. The two force arrows will be at a 45° angle to one another.
   d. The two force arrows will be at a 180° angle to one another.

3. A free-body diagram shows the forces acting on an object. How is that object represented in the diagram?
   a. A single point
   b. A square box
   c. A unit circle
   d. The object as it is

4.2 Newton's First Law of Motion: Inertia

4. A ball rolls along the ground, moving from north to south. What direction is the frictional force that acts on the ball?
   a. North to south
   b. South to north
   c. West to east
   d. East to west

5. The tires you choose to drive over icy roads will create more friction with the road than your summer tires. Give another example where more friction is desirable.

a. Children's slide
b. Air hockey table
c. Ice-skating rink
d. Jogging track

6. How do you express, mathematically, that no external force is acting on a body?
   a. $F_{net} = -1$
   b. $F_{net} = 0$
   c. $F_{net} = 1$
   d. $F_{net} = \infty$

4.3 Newton's Second Law of Motion

7. What does it mean for two quantities to be inversely proportional to each other?
   a. When one variable increases, the other variable decreases by a greater amount.
   b. When one variable increases, the other variable also increases.
   c. When one variable increases, the other variable decreases by the same factor.
   d. When one variable increases, the other variable also increases by the same factor.

8. True or False: Newton's second law can be interpreted based on Newton's first law.
   a. True
   b. False

4.4 Newton's Third Law of Motion

9. Which forces cause changes in the motion of a system?
   a. internal forces
   b. external forces
   c. both internal and external forces
   d. neither internal nor external forces

10. True or False—Newton's third law applies to the external forces acting on a system of interest.
   a. True
   b. False

11. A ball is dropped and hits the floor. What is the direction of the force exerted by the floor on the ball?
   a. Upward
   b. Downward
   c. Right
   d. Left

Critical Thinking Items

4.1 Force

12. Only two forces are acting on an object: force A to the left and force B to the right. If force B is greater than force A, in which direction will the object move?
   a. To the right
   b. To the left
   c. Upward
   d. The object does not move

13. In a free-body diagram, the arrows representing tension and weight have the same length but point away from one another. What does this indicate?
   a. They are equal in magnitude and act in the same direction.
   b. They are equal in magnitude and act in opposite directions.
   c. They are unequal in magnitude and act in the same direction.
   d. They are unequal in magnitude and act in opposite directions.

14. An object is at rest. Two forces, X and Y, are acting on it. Force X has a magnitude of x and acts in the downward direction. What is the magnitude and direction of Y?
   a. The magnitude is x and points in the upward direction.
   b. The magnitude is $2x$ and points in the upward direction.
   c. The magnitude is x and points in the downward direction.
   d. The magnitude is $2x$ and points in the downward direction.

15. Three forces, A, B, and C, are acting on the same object with magnitudes a, b, and c, respectively. Force A acts to the right, force B acts to the left, and force C acts downward. What is a necessary condition for the object to move straight down?
   a. The magnitude of force A must be greater than the magnitude of force B, so $a > b$.
   b. The magnitude of force A must be equal to the magnitude of force B, so $a = b$.
   c. The magnitude of force A must be greater than the magnitude of force C, so $A > C$.
   d. The magnitude of force C must be greater than the magnitude of forces A or B, so $A < C > B$.

4.2 Newton's First Law of Motion: Inertia

16. Two people push a cart on a horizontal surface by applying forces F_1 and F_2 in the same direction. Is the magnitude of the net force acting on the cart, F_{net}, equal to, greater than, or less than $F_1 + F_2$? Why?
   a. $F_{net} < F_1 + F_2$ because the net force will not include the frictional force.
   b. $F_{net} = F_1 + F_2$ because the net force will not include

the frictional force

c. $F_{net} < F_1 + F_2$ because the net force will include the component of frictional force

d. $F_{net} = F_1 + F_2$ because the net force will include the frictional force

17. True or False: A book placed on a balance scale is balanced by a standard 1-kg iron weight placed on the opposite side of the balance. If these objects are taken to the moon and a similar exercise is performed, the balance is still level because gravity is uniform on the moon's surface as it is on Earth's surface.

a. True

b. False

4.3 Newton's Second Law of Motion

18. From the equation for Newton's second law, we see that F_{net} is directly proportional to **a** and that the constant of proportionality is ***m***. What does this mean in a practical sense?

a. An increase in applied force will cause an increase in acceleration if the mass is constant.

b. An increase in applied force will cause a decrease in acceleration if the mass is constant.

c. An increase in applied force will cause an increase in acceleration, even if the mass varies.

d. An increase in applied force will cause an increase in acceleration and mass.

4.4 Newton's Third Law of Motion

19. True or False: A person accelerates while walking on the ground by exerting force. The ground in turn exerts force F_2 on the person. F_1 and F_2 are equal in magnitude but act in opposite directions. The person is able to walk because the two forces act on the different systems and the net force acting on the person is nonzero.

a. True

b. False

20. A helicopter pushes air down, which, in turn, pushes the helicopter up. Which force affects the helicopter's motion? Why?

a. Air pushing upward affects the helicopter's motion because it is an internal force that acts on the helicopter.

b. Air pushing upward affects the helicopter's motion because it is an external force that acts on the helicopter.

c. The downward force applied by the blades of the helicopter affects its motion because it is an internal force that acts on the helicopter.

d. The downward force applied by the blades of the helicopter affects its motion because it is an external force that acts on the helicopter.

Problems

4.3 Newton's Second Law of Motion

21. An object has a mass of 1 kg on Earth. What is its weight on the moon?

a. 1 N

b. 1.67 N

c. 9.8 N

d. 10 N

22. A bathroom scale shows your mass as 55 kg. What will it read on the moon?

a. 9.4 kg

b. 10.5 kg

c. 55.0 kg

d. 91.9 kg

4.4 Newton's Third Law of Motion

23. A person pushes an object of mass 5.0 kg along the floor by applying a force. If the object experiences a friction force of 10 N and accelerates at 18 m/s^2, what is the magnitude of the force exerted by the person?

a. −90 N

b. −80 N

c. 90 N

d. 100 N

Performance Task

4.4 Newton's Third Law of Motion

24. A car weighs 2,000 kg. It moves along a road by applying a force on the road with a parallel component of 560 N. There are two passengers in the car, each weighing 55 kg. If the magnitude of the force of friction experienced by the car is 45 N, what is the acceleration of the car?

a. 0.244 m/s^2

b. 0.265 m/s^2

c. 4.00 m/s^2

d. 4.10 m/s^2

TEST PREP

Multiple Choice

4.1 Force

25. Which of the following is a physical quantity that can be described by dynamics but not by kinematics?
a. Velocity
b. Acceleration
c. Force

26. Which of the following is used to represent an object in a free-body diagram?
a. A point
b. A line
c. A vector

4.2 Newton's First Law of Motion: Inertia

27. What kind of force is friction?
a. External force
b. Internal force
c. Net force

28. What is another name for Newton's first law?
a. Law of infinite motion
b. Law of inertia
c. Law of friction

29. True or False—A rocket is launched into space and escapes Earth's gravitational pull. It will continue to travel in a straight line until it is acted on by another force.
a. True
b. False

30. A 2,000-kg car is sitting at rest in a parking lot. A bike and rider with a total mass of 60 kg are traveling along a road at 10 km/h. Which system has more inertia? Why?
a. The car has more inertia, as its mass is greater than the mass of the bike.
b. The bike has more inertia, as its mass is greater than the mass of the car.
c. The car has more inertia, as its mass is less than the mass of the bike.
d. The bike has more inertia, as its mass is less than the mass of the car.

4.3 Newton's Second Law of Motion

31. In the equation for Newton's second law, what does F_{net} stand for?
a. Internal force
b. Net external force
c. Frictional force

32. What is the SI unit of force?
a. Kg
b. dyn
c. N

33. What is the net external force on an object in freefall on Earth if you were to neglect the effect of air?
a. The net force is zero.
b. The net force is upward with magnitude *mg*.
c. The net force is downward with magnitude *mg*.
d. The net force is downward with magnitude 9.8 N.

34. Two people push a 2,000-kg car to get it started. An acceleration of at least 5.0 m/s^2 is required to start the car. Assuming both people apply the same magnitude force, how much force will each need to apply if friction between the car and the road is 300 N?
a. 4850 N
b. 5150 N
c. 97000 N
d. 10300 N

4.4 Newton's Third Law of Motion

35. One object exerts a force of magnitude F_1 on another object and experiences a force of magnitude F_2 in return. What is true for F_1 and F_2?
a. $F_1 > F_2$
b. $F_1 < F_2$
c. $F_1 = F_2$

36. A weight is suspended with a rope and hangs freely. In what direction is the tension on the rope?
a. parallel to the rope
b. perpendicular to the rope

37. A person weighing 55 kg walks by applying 160 N of force on the ground, while pushing a 10-kg object. If the person accelerates at 2 m/s^2, what is the force of friction experienced by the system consisting of the person and the object?
a. 30 N
b. 50 N
c. 270 N
d. 290 N

38. A 65-kg swimmer pushes on the pool wall and accelerates at 6 m/s^2. The friction experienced by the swimmer is 100 N. What is the magnitude of the force that the swimmer applies on the wall?
a. −490 N
b. −290 N
c. 290 N
d. 490 N

Short Answer

4.1 Force

39. True or False—An external force is defined as a force generated outside the system of interest that acts on an object inside the system.
   a. True
   b. False

40. By convention, which sign is assigned to an object moving downward?
   a. A positive sign (+)
   b. A negative sign (−)
   c. Either a positive or negative sign (±)
   d. No sign is assigned

41. A body is pushed downward by a force of 5 units and upward by a force of 2 units. How would you draw a free-body diagram to represent this?
   a. Two force vectors acting at a point, both pointing up with lengths of 5 units and 2 units
   b. Two force vectors acting at a point, both pointing down with lengths of 5 units and 2 units
   c. Two force vectors acting at a point, one pointing up with a length of 5 units and the other pointing down with a length of 2 units
   d. Two force vectors acting at a point, one pointing down with a length of 5 units and the other pointing up with a length of 2 units

42. A body is pushed eastward by a force of four units and southward by a force of three units. How would you draw a free-body diagram to represent this?
   a. Two force vectors acting at a point, one pointing left with a length of 4 units and the other pointing down with a length of 3 units
   b. Two force vectors acting at a point, one pointing left with a length of 4 units and the other pointing up with a length of 3 units
   c. Two force vectors acting at a point, one pointing right with a length of 4 units and the other pointing down with a length of 3 units
   d. Two force vectors acting at a point, one pointing right with a length of 4 units and the other pointing up with a length of 3 units

4.2 Newton's First Law of Motion: Inertia

43. A body with mass m is pushed along a horizontal surface by a force F and is opposed by a frictional force *f*. How would you draw a free-body diagram to represent this situation?
   a. A dot with an arrow pointing right, labeled F, and an arrow pointing left, labeled *f*, that is of equal length or shorter than F
   b. A dot with an arrow pointing right, labeled F, and an arrow pointing right, labeled *f*, that is of equal length or shorter than F
   c. A dot with an arrow pointing right, labeled F, and a smaller arrow pointing up, labeled *f*, that is of equal length or longer than F
   d. A dot with an arrow pointing right, labeled F, and a smaller arrow pointing down, labeled *f*, that is of equal length or longer than F

44. Two objects rest on a uniform surface. A person pushes both with equal force. If the first object starts to move faster than the second, what can be said about their masses?
   a. The mass of the first object is less than that of the second object.
   b. The mass of the first object is equal to the mass of the second object.
   c. The mass of the first object is greater than that of the second object.
   d. No inference can be made because mass and force are not related to each other.

45. Two similar boxes rest on a table. One is empty and the other is filled with pebbles. Without opening or lifting either, how can you tell which box is full? Why?
   a. By applying an internal force; whichever box accelerates faster is lighter and so must be empty
   b. By applying an internal force; whichever box accelerates faster is heavier and so the other box must be empty
   c. By applying an external force; whichever box accelerates faster is lighter and so must be empty
   d. By applying an external force; whichever box accelerates faster is heavier and so the other box must be empty

46. True or False—An external force is required to set a stationary object in motion in outer space away from all gravitational influences and atmospheric friction.
   a. True
   b. False

4.3 Newton's Second Law of Motion

47. A steadily rolling ball is pushed in the direction from east to west, which causes the ball to move faster in the same direction. What is the direction of the acceleration?
   a. North to south
   b. South to north
   c. East to west
   d. West to east

48. A ball travels from north to south at 60 km/h. After being hit by a bat, it travels from west to east at 60 km/

h. Is there a change in velocity?
   a. Yes, because velocity is a scalar.
   b. Yes, because velocity is a vector.
   c. No, because velocity is a scalar.
   d. No, because velocity is a vector

49. What is the weight of a 5-kg object on Earth and on the moon?
   a. On Earth the weight is 1.67 N, and on the moon the weight is 1.67 N.
   b. On Earth the weight is 5 N, and on the moon the weight is 5 N.
   c. On Earth the weight is 49 N, and on the moon the weight is 8.35 N.
   d. On Earth the weight is 8.35 N, and on the moon the weight is 49 N.

50. An object weighs 294 N on Earth. What is its weight on the moon?
   a. 50.1 N
   b. 30.0 N
   c. 249 N
   d. 1461 N

4.4 Newton's Third Law of Motion

51. A large truck with mass 30 m crashes into a small sedan with mass *m*. If the truck exerts a force F on the sedan, what force will the sedan exert on the truck?
   a. $\frac{F}{30}$
   b. F
   c. 2F
   d. 30F

52. A fish pushes water backward with its fins. How does this propel the fish forward?
   a. The water exerts an internal force on the fish in the opposite direction, pushing the fish forward.
   b. The water exerts an external force on the fish in the opposite direction, pushing the fish forward.
   c. The water exerts an internal force on the fish in the same direction, pushing the fish forward.
   d. The water exerts an external force on the fish in the same direction, pushing the fish forward.

53. True or False—Tension is the result of opposite forces in a connector, such as a string, rope, chain or cable, that pulls each point of the connector apart in the direction parallel to the length of the connector. At the ends of the connector, the tension pulls toward the center of the connector.
   a. True
   b. False

54. True or False—Normal reaction is the force that opposes the force of gravity and acts in the direction of the force of gravity.
   a. True
   b. False

Extended Response

4.1 Force

55. True or False—When two unequal forces act on a body, the body will not move in the direction of the weaker force.
   a. True
   b. False

56. In the figure given, what is $F_{restore}$? What is its magnitude?

   a. $F_{restore}$ is the force exerted by the hand on the spring, and it pulls to the right.
   b. $F_{restore}$ is the force exerted by the spring on the hand, and it pulls to the left.
   c. $F_{restore}$ is the force exerted by the hand on the spring, and it pulls to the left.
   d. $F_{restore}$ is the force exerted by the spring on the hand, and it pulls to the right.

4.2 Newton's First Law of Motion: Inertia

57. Two people apply the same force to throw two identical balls in the air. Will the balls necessarily travel the same distance? Why or why not?
   a. No, the balls will not necessarily travel the same distance because the gravitational force acting on them is different.
   b. No, the balls will not necessarily travel the same distance because the angle at which they are thrown may differ.
   c. Yes, the balls will travel the same distance because the gravitational force acting on them is the same.
   d. Yes, the balls will travel the same distance because the angle at which they are thrown may differ.

58. A person pushes a box from left to right and then lets the box slide freely across the floor. The box slows down as it slides across the floor. When the box is sliding

freely, what is the direction of the net external force?

a. The net external force acts from left to right.
b. The net external force acts from right to left.
c. The net external force acts upward.
d. The net external force acts downward.

4.3 Newton's Second Law of Motion

59. A 55-kg lady stands on a bathroom scale inside an elevator. The scale reads 70 kg. What do you know about the motion of the elevator?

a. The elevator must be accelerating upward.
b. The elevator must be accelerating downward.
c. The elevator must be moving upward with a constant velocity.
d. The elevator must be moving downward with a constant velocity.

60. True or False—A skydiver initially accelerates in his jump. Later, he achieves a state of constant velocity called terminal velocity. Does this mean the skydiver becomes weightless?

a. Yes
b. No

4.4 Newton's Third Law of Motion

61. How do rockets propel themselves in space?

a. Rockets expel gas in the forward direction at high velocity, and the gas, which provides an internal force, pushes the rockets forward.
b. Rockets expel gas in the forward direction at high velocity, and the gas, which provides an external force, pushes the rockets forward.
c. Rockets expel gas in the backward direction at high velocity, and the gas, which is an internal force, pushes the rockets forward.
d. Rockets expel gas in the backward direction at high velocity, and the gas, which provides an external force, pushes the rockets forward.

62. Are rockets more efficient in Earth's atmosphere or in outer space? Why?

a. Rockets are more efficient in Earth's atmosphere than in outer space because the air in Earth's atmosphere helps to provide thrust for the rocket, and Earth has more air friction than outer space.
b. Rockets are more efficient in Earth's atmosphere than in outer space because the air in Earth's atmosphere helps to provide thrust to the rocket, and Earth has less air friction than the outer space.
c. Rockets are more efficient in outer space than in Earth's atmosphere because the air in Earth's atmosphere does not provide thrust but does create more air friction than in outer space.
d. Rockets are more efficient in outer space than in Earth's atmosphere because the air in Earth's atmosphere does not provide thrust but does create less air friction than in outer space.

CHAPTER 5
Motion in Two Dimensions

Figure 5.1 Billiard balls on a pool table are in motion after being hit with a cue stick. (Popperipopp, Wikimedia Commons)

Chapter Outline

5.1 Vector Addition and Subtraction: Graphical Methods

5.2 Vector Addition and Subtraction: Analytical Methods

5.3 Projectile Motion

5.4 Inclined Planes

5.5 Simple Harmonic Motion

INTRODUCTION In Chapter 2, we learned to distinguish between vectors and scalars; the difference being that a vector has magnitude and direction, whereas a scalar has only magnitude. We learned how to deal with vectors in physics by working straightforward one-dimensional vector problems, which may be treated mathematically in the same as scalars. In this chapter, we'll use vectors to expand our understanding of forces and motion into two dimensions. Most real-world physics problems (such as with the game of pool pictured here) are, after all, either two- or three-dimensional problems and physics is most useful when applied to real physical scenarios. We start by learning the practical skills of graphically adding and subtracting vectors (by using drawings) and analytically (with math). Once we're able to work with two-dimensional vectors, we apply these skills to problems of projectile motion, inclined planes, and harmonic motion.

5.1 Vector Addition and Subtraction: Graphical Methods

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Describe the graphical method of vector addition and subtraction
- Use the graphical method of vector addition and subtraction to solve physics problems

Section Key Terms

graphical method	head (of a vector)	head-to-tail method	resultant
resultant vector	tail	vector addition	vector subtraction

The Graphical Method of Vector Addition and Subtraction

Recall that a vector is a quantity that has magnitude and direction. For example, displacement, velocity, acceleration, and force are all vectors. In one-dimensional or straight-line motion, the direction of a vector can be given simply by a plus or minus sign. Motion that is forward, to the right, or upward is usually considered to be *positive* (+); and motion that is backward, to the left, or downward is usually considered to be *negative* (–).

In two dimensions, a vector describes motion in two perpendicular directions, such as vertical and horizontal. For vertical and horizontal motion, each vector is made up of vertical and horizontal components. In a one-dimensional problem, one of the components simply has a value of zero. For two-dimensional vectors, we work with vectors by using a frame of reference such as a coordinate system. Just as with one-dimensional vectors, we graphically represent vectors with an arrow having a length proportional to the vector's magnitude and pointing in the direction that the vector points.

Figure 5.2 shows a graphical representation of a vector; the total displacement for a person walking in a city. The person first walks nine blocks east and then five blocks north. Her total displacement does not match her path to her final destination. The displacement simply connects her starting point with her ending point using a straight line, which is the shortest distance. We use the notation that a boldface symbol, such as **D**, stands for a vector. Its magnitude is represented by the symbol in italics, D, and its direction is given by an angle represented by the symbol θ. Note that her displacement would be the same if she had begun by first walking five blocks north and then walking nine blocks east.

TIPS FOR SUCCESS

In this text, we represent a vector with a boldface variable. For example, we represent a force with the vector **F**, which has both magnitude and direction. The magnitude of the vector is represented by the variable in italics, F, and the direction of the variable is given by the angle θ.

Figure 5.2 A person walks nine blocks east and five blocks north. The displacement is 10.3 blocks at an angle 29.1° north of east.

The **head-to-tail method** is a **graphical** way to add vectors. The **tail** of the vector is the starting point of the vector, and the **head** (or tip) of a vector is the pointed end of the arrow. The following steps describe how to use the head-to-tail method for graphical **vector addition**.

1. Let the *x*-axis represent the east-west direction. Using a ruler and protractor, draw an arrow to represent the first vector (nine blocks to the east), as shown in Figure 5.3(a).

Figure 5.3 The diagram shows a vector with a magnitude of nine units and a direction of 0°.

2. Let the *y*-axis represent the north-south direction. Draw an arrow to represent the second vector (five blocks to the north). Place the tail of the second vector at the head of the first vector, as shown in Figure 5.4(b).

Figure 5.4 A vertical vector is added.

3. If there are more than two vectors, continue to add the vectors head-to-tail as described in step 2. In this example, we have only two vectors, so we have finished placing arrows tip to tail.
4. Draw an arrow from the tail of the first vector to the head of the last vector, as shown in Figure 5.5(c). This is the **resultant**, or the sum, of the vectors.

Figure 5.5 The diagram shows the resultant vector, a ruler, and protractor.

5. To find the magnitude of the resultant, measure its length with a ruler. When we deal with vectors analytically in the next section, the magnitude will be calculated by using the Pythagorean theorem.
6. To find the direction of the resultant, use a protractor to measure the angle it makes with the reference direction (in this case, the *x*-axis). When we deal with vectors analytically in the next section, the direction will be calculated by using trigonometry to find the angle.

WATCH PHYSICS

Visualizing Vector Addition Examples

This video shows four graphical representations of vector addition and matches them to the correct vector addition formula.

Click to view content (https://openstax.org/l/02addvector)

GRASP CHECK

There are two vectors $\vec{a}$ and $\vec{b}$. The head of vector $\vec{a}$ touches the tail of vector $\vec{b}$. The addition of vectors $\vec{a}$ and $\vec{b}$ gives a resultant vector $\vec{c}$. Can the addition of these two vectors can be represented by the following two equations? $\vec{a} + \vec{b} = \vec{c}$; $\vec{b} + \vec{a} = \vec{c}$

a. Yes, if we add the same two vectors in a different order it will still give the same resultant vector.
b. No, the resultant vector will change if we add the same vectors in a different order.

Vector subtraction is done in the same way as vector addition with one small change. We add the first vector to the negative of the vector that needs to be subtracted. A negative vector has the same magnitude as the original vector, but points in the opposite direction (as shown in Figure 5.6). Subtracting the vector **B** from the vector **A**, which is written as **A** – **B**, is the same as **A** + (–**B**). Since it does not matter in what order vectors are added, **A** – **B** is also equal to (–**B**) + **A**. This is true for scalars as well as vectors. For example, 5 – 2 = 5 + (–2) = (–2) + 5.

Figure 5.6 The diagram shows a vector, B, and the negative of this vector, −B.

Global angles are calculated in the counterclockwise direction. The clockwise direction is considered negative. For example, an angle of 30° south of west is the same as the global angle 210°, which can also be expressed as -150° from the positive *x*-axis.

Using the Graphical Method of Vector Addition and Subtraction to Solve Physics Problems

Now that we have the skills to work with vectors in two dimensions, we can apply vector addition to graphically determine the **resultant vector**, which represents the total force. Consider an example of force involving two ice skaters pushing a third as seen in Figure 5.7.

Figure 5.7 Part (a) shows an overhead view of two ice skaters pushing on a third. Forces are vectors and add like vectors, so the total force on the third skater is in the direction shown. In part (b), we see a free-body diagram representing the forces acting on the third skater.

In problems where variables such as force are already known, the forces can be represented by making the length of the vectors proportional to the magnitudes of the forces. For this, you need to create a scale. For example, each centimeter of vector length could represent 50 N worth of force. Once you have the initial vectors drawn to scale, you can then use the head-to-tail method to draw the resultant vector. The length of the resultant can then be measured and converted back to the original units using the scale you created.

You can tell by looking at the vectors in the free-body diagram in Figure 5.7 that the two skaters are pushing on the third skater with equal-magnitude forces, since the length of their force vectors are the same. Note, however, that the forces are not equal because they act in different directions. If, for example, each force had a magnitude of 400 N, then we would find the magnitude of the total external force acting on the third skater by finding the magnitude of the resultant vector. Since the forces act at a right angle to one another, we can use the Pythagorean theorem. For a triangle with sides a, b, and c, the Pythagorean theorem tells us that

$$a^2 + b^2 = c^2$$
$$c = \sqrt{a^2 + b^2}.$$

Applying this theorem to the triangle made by $\mathbf{F}_1$, $\mathbf{F}_2$, and $\mathbf{F}_{tot}$ in Figure 5.7, we get

$$\mathbf{F}_{tot}^2 = \sqrt{\mathbf{F}_1^2 + \mathbf{F}_1^2},$$

or

$$\mathbf{F}_{\text{tot}} = \sqrt{(400\text{ N})^2 + (400\text{ N})^2} = 566\text{ N}.$$

Note that, if the vectors were not at a right angle to each other (90° to one another), we would not be able to use the Pythagorean theorem to find the magnitude of the resultant vector. Another scenario where adding two-dimensional vectors is necessary is for velocity, where the direction may not be purely east-west or north-south, but some combination of these two directions. In the next section, we cover how to solve this type of problem analytically. For now let's consider the problem graphically.

WORKED EXAMPLE

Adding Vectors Graphically by Using the Head-to-Tail Method: A Woman Takes a Walk

Use the graphical technique for adding vectors to find the total displacement of a person who walks the following three paths (displacements) on a flat field. First, he walks 25 m in a direction 49° north of east. Then, he walks 23 m heading 15° north of east. Finally, he turns and walks 32 m in a direction 68° south of east.

Strategy

Graphically represent each displacement vector with an arrow, labeling the first **A**, the second **B**, and the third **C**. Make the lengths proportional to the distance of the given displacement and orient the arrows as specified relative to an east-west line. Use the head-to-tail method outlined above to determine the magnitude and direction of the resultant displacement, which we'll call **R**.

Solution

(1) Draw the three displacement vectors, creating a convenient scale (such as 1 cm of vector length on paper equals 1 m in the problem), as shown in Figure 5.8.

Figure 5.8 The three displacement vectors are drawn first.

(2) Place the vectors head to tail, making sure not to change their magnitude or direction, as shown in Figure 5.9.

Figure 5.9 Next, the vectors are placed head to tail.

(3) Draw the **resultant vector R** from the tail of the first vector to the head of the last vector, as shown in Figure 5.10.

Figure 5.10 The resultant vector is drawn .

(4) Use a ruler to measure the magnitude of **R**, remembering to convert back to the units of meters using the scale. Use a protractor to measure the direction of **R**. While the direction of the vector can be specified in many ways, the easiest way is to measure the angle between the vector and the nearest horizontal or vertical axis. Since **R** is south of the eastward pointing axis (the *x*-axis), we flip the protractor upside down and measure the angle between the eastward axis and the vector, as illustrated in Figure 5.11.

Figure 5.11 A ruler is used to measure the magnitude of **R**, and a protractor is used to measure the direction of **R**.

In this case, the total displacement **R** has a magnitude of 50 m and points 7° south of east. Using its magnitude and direction, this vector can be expressed as

$$\mathrm{R} = 50 \ \ \mathrm{m} \tag{5.1}$$

and

$$\theta = 7^{\circ}\text{south of east.} \tag{5.2}$$

Discussion

The head-to-tail graphical method of vector addition works for any number of vectors. It is also important to note that it does not matter in what order the vectors are added. Changing the order does not change the resultant. For example, we could add the vectors as shown in Figure 5.12, and we would still get the same solution.

Figure 5.12 Vectors can be added in any order to get the same result.

WORKED EXAMPLE

Subtracting Vectors Graphically: A Woman Sailing a Boat

A woman sailing a boat at night is following directions to a dock. The instructions read to first sail 27.5 m in a direction 66.0° north of east from her current location, and then travel 30.0 m in a direction 112° north of east (or 22.0° west of north). If the woman makes a mistake and travels in the *opposite* direction for the second leg of the trip, where will she end up? The two legs of the woman's trip are illustrated in Figure 5.13.

Figure 5.13 In the diagram, the first leg of the trip is represented by vector A and the second leg is represented by vector B.

Strategy

We can represent the first leg of the trip with a vector **A**, and the second leg of the trip that she was *supposed to* take with a vector **B**. Since the woman mistakenly travels in the opposite direction for the second leg of the journey, the vector for second leg of the trip she *actually* takes is –**B**. Therefore, she will end up at a location **A** + (–**B**), or **A** – **B**. Note that –**B** has the same magnitude as **B** (30.0 m), but is in the opposite direction, $68^\circ (180^\circ - 112^\circ)$ south of east, as illustrated in Figure 5.14.

Figure 5.14 Vector –B represents traveling in the opposite direction of vector B.

We use graphical vector addition to find where the woman arrives **A** + (–**B**).

Solution

(1) To determine the location at which the woman arrives by accident, draw vectors **A** and –**B**.

(2) Place the vectors head to tail.

(3) Draw the resultant vector **R**.

(4) Use a ruler and protractor to measure the magnitude and direction of **R**.

These steps are demonstrated in Figure 5.15.

Figure 5.15 The vectors are placed head to tail.

In this case

$$R = 23.0 \text{ m} \tag{5.3}$$

and

$$\theta = 7.5° \text{south of east.} \tag{5.4}$$

Discussion

Because subtraction of a vector is the same as addition of the same vector with the opposite direction, the graphical method for subtracting vectors works the same as for adding vectors.

WORKED EXAMPLE

Adding Velocities: A Boat on a River

A boat attempts to travel straight across a river at a speed of 3.8 m/s. The river current flows at a speed v_{river} of 6.1 m/s to the right. What is the total velocity and direction of the boat? You can represent each meter per second of velocity as one centimeter of vector length in your drawing.

Strategy

We start by choosing a coordinate system with its x-axis parallel to the velocity of the river. Because the boat is directed straight toward the other shore, its velocity is perpendicular to the velocity of the river. We draw the two vectors, $\mathbf{v}_{\text{boat}}$ and $\mathbf{v}_{\text{river}}$, as shown in Figure 5.16.

Using the head-to-tail method, we draw the resulting total velocity vector from the tail of $\mathbf{v}_{\text{boat}}$ to the head of $\mathbf{v}_{\text{river}}$.

Figure 5.16 A boat attempts to travel across a river. What is the total velocity and direction of the boat?

Solution

By using a ruler, we find that the length of the resultant vector is 7.2 cm, which means that the magnitude of the total velocity is

$$v_{tot} = 7.2 \text{ m/s}. \tag{5.5}$$

By using a protractor to measure the angle, we find $\theta = 32.0^\circ$.

Discussion

If the velocity of the boat and river were equal, then the direction of the total velocity would have been 45°. However, since the velocity of the river is greater than that of the boat, the direction is less than 45° with respect to the shore, or *x* axis.

Practice Problems

1. Vector $\vec{A}$, having magnitude 2.5 m, pointing 37° south of east and vector $\vec{B}$ having magnitude 3.5 m, pointing 20° north of east are added. What is the magnitude of the resultant vector?
   a. 1.0 m
   b. 5.3 m
   c. 5.9 m
   d. 6.0 m

2. A person walks 32° north of west for 94 m and 35° east of south for 122 m. What is the magnitude of his displacement?
   a. 28 m
   b. 51 m
   c. 180 m
   d. 216 m

Virtual Physics

Vector Addition

In this simulation (https://archive.cnx.org/specials/d218bf9b-e50e-4d50-9a6c-b3db4dad0816/vector-addition/) , you will experiment with adding vectors graphically. Click and drag the red vectors from the Grab One basket onto the graph in the middle of the screen. These red vectors can be rotated, stretched, or repositioned by clicking and dragging with your mouse. Check the Show Sum box to display the resultant vector (in green), which is the sum of all of the red vectors placed on the

graph. To remove a red vector, drag it to the trash or click the Clear All button if you wish to start over. Notice that, if you click on any of the vectors, the $|\mathbf{R}|$ is its magnitude, θ is its direction with respect to the positive x-axis, $\mathbf{R_x}$ is its horizontal component, and R_y is its vertical component. You can check the resultant by lining up the vectors so that the head of the first vector touches the tail of the second. Continue until all of the vectors are aligned together head-to-tail. You will see that the resultant magnitude and angle is the same as the arrow drawn from the tail of the first vector to the head of the last vector. Rearrange the vectors in any order head-to-tail and compare. The resultant will always be the same.

Click to view content (https://archive.cnx.org/specials/d218bf9b-e50e-4d50-9a6c-b3db4dad0816/vector-addition/)

GRASP CHECK

True or False—The more long, red vectors you put on the graph, rotated in any direction, the greater the magnitude of the resultant green vector.

a. True
b. False

Check Your Understanding

3. While there is no single correct choice for the sign of axes, which of the following are conventionally considered positive?
   a. backward and to the left
   b. backward and to the right
   c. forward and to the right
   d. forward and to the left

4. True or False—A person walks 2 blocks east and 5 blocks north. Another person walks 5 blocks north and then two blocks east. The displacement of the first person will be more than the displacement of the second person.
   a. True
   b. False

5.2 Vector Addition and Subtraction: Analytical Methods

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Define components of vectors
- Describe the analytical method of vector addition and subtraction
- Use the analytical method of vector addition and subtraction to solve problems

Section Key Terms

analytical method	component (of a two-dimensional vector)

Components of Vectors

For the **analytical method** of vector addition and subtraction, we use some simple geometry and trigonometry, instead of using a ruler and protractor as we did for graphical methods. However, the graphical method will still come in handy to visualize the problem by drawing vectors using the head-to-tail method. The analytical method is more accurate than the graphical method, which is limited by the precision of the drawing. For a refresher on the definitions of the sine, cosine, and tangent of an angle, see Figure 5.17.

$$\sin(\theta) = \frac{y}{h}$$
$$\cos(\theta) = \frac{x}{h}$$
$$\tan(\theta) = \frac{y}{x}$$

Figure 5.17 For a right triangle, the sine, cosine, and tangent of θ are defined in terms of the adjacent side, the opposite side, or the hypotenuse. In this figure, x is the adjacent side, y is the opposite side, and h is the hypotenuse.

Since, by definition, $\cos\theta = x/h$, we can find the length x if we know h and θ by using $x = h\cos\theta$. Similarly, we can find the length of y by using $y = h\sin\theta$. These trigonometric relationships are useful for adding vectors.

When a vector acts in more than one dimension, it is useful to break it down into its x and y components. For a two-dimensional vector, a **component** is a piece of a vector that points in either the x- or y-direction. Every 2-d vector can be expressed as a sum of its x and y components.

For example, given a vector like $\mathbf{A}$ in Figure 5.18, we may want to find what two perpendicular vectors, $\mathbf{A}_x$ and $\mathbf{A}_y$, add to produce it. In this example, $\mathbf{A}_x$ and $\mathbf{A}_y$ form a right triangle, meaning that the angle between them is 90 degrees. This is a common situation in physics and happens to be the least complicated situation trigonometrically.

Figure 5.18 The vector $\mathbf{A}$, with its tail at the origin of an x- y-coordinate system, is shown together with its x- and y-components, $\mathbf{A}_x$ and $\mathbf{A}_y$. These vectors form a right triangle.

$\mathbf{A}_x$ and $\mathbf{A}_y$ are defined to be the components of $\mathbf{A}$ along the x- and y-axes. The three vectors, $\mathbf{A}$, $\mathbf{A}_x$, and $\mathbf{A}_y$, form a right triangle.

$$\mathbf{A_x} + \mathbf{A_y} = \mathbf{A}$$

If the vector $\mathbf{A}$ is known, then its magnitude A (its length) and its angle θ (its direction) are known. To find A_x and A_y, its x- and y-components, we use the following relationships for a right triangle:

$$A_x = A\cos\theta$$

and

$$A_y = A\sin\theta,$$

where A_x is the magnitude of $\mathbf{A}$ in the x-direction, A_y is the magnitude of $\mathbf{A}$ in the y-direction, and θ is the angle of the resultant with respect to the x-axis, as shown in Figure 5.19.

Figure 5.19 The magnitudes of the vector components $\mathbf{A}_x$ and $\mathbf{A}_y$ can be related to the resultant vector $\mathbf{A}$ and the angle θ with trigonometric identities. Here we see that $A_x = A\cos\theta$ and $A_y = A\sin\theta$.

Suppose, for example, that $\mathbf{A}$ is the vector representing the total displacement of the person walking in a city, as illustrated in Figure 5.20.

Figure 5.20 We can use the relationships $A_x = A\cos\theta$ and $A_y = A\sin\theta$ to determine the magnitude of the horizontal and vertical component vectors in this example.

Then A = 10.3 blocks and $\theta = 29.1^\circ$, so that

$$\begin{aligned} A_x &= A\cos\theta \\ &= (10.3\ \text{blocks})(\cos 29.1^\circ) \\ &= (10.3\ \text{blocks})(0.874) \\ &= 9.0\ \text{blocks}. \end{aligned} \tag{5.6}$$

This magnitude indicates that the walker has traveled 9 blocks to the east—in other words, a 9-block eastward displacement. Similarly,

$$\begin{aligned} A_y &= A\sin\theta \\ &= (10.3\ \text{blocks})(\sin 29.1^\circ) \\ &= (10.3\ \text{blocks})(0.846) \\ &= 5.0\ \text{blocks}, \end{aligned} \tag{5.7}$$

indicating that the walker has traveled 5 blocks to the north—a 5-block northward displacement.

Analytical Method of Vector Addition and Subtraction

Calculating a resultant vector (or vector addition) is the reverse of breaking the resultant down into its components. If the perpendicular components $\mathbf{A}_x$ and $\mathbf{A}_y$ of a vector $\mathbf{A}$ are known, then we can find $\mathbf{A}$ analytically. How do we do this? Since, by definition,

$$\tan\theta = y/x \text{ (or in this case } \tan\theta = A_y/A_x\text{)},$$

we solve for θ to find the direction of the resultant.

$$\theta = \tan^{-1}(A_y/A_x)$$

Since this is a right triangle, the Pythagorean theorem ($x^2 + y^2 = h^2$) for finding the hypotenuse applies. In this case, it becomes

$$A^2 = A_x^2 + A_y^2.$$

Solving for A gives

$$A = \sqrt{{A_x}^2 + {A_y}^2}.$$

In summary, to find the magnitude A and direction θ of a vector from its perpendicular components $\mathbf{A}_x$ and $\mathbf{A}_y$, as illustrated in Figure 5.21, we use the following relationships:

$$\begin{aligned} A &= \sqrt{{A_x}^2 + {A_y}^2} \\ \theta &= \tan^{-1}(A_y/A_x) \end{aligned}$$

Figure 5.21 The magnitude and direction of the resultant vector $\mathbf{A}$ can be determined once the horizontal components $\mathbf{A}_x$ and $\mathbf{A}_y$ have been determined.

Sometimes, the vectors added are not perfectly perpendicular to one another. An example of this is the case below, where the vectors $\mathbf{A}$ and $\mathbf{B}$ are added to produce the resultant $\mathbf{R}$, as illustrated in Figure 5.22.

Figure 5.22 Vectors **A** and **B** are two legs of a walk, and **R** is the resultant or total displacement. You can use analytical methods to determine the magnitude and direction of **R** .

If **A** and **B** represent two legs of a walk (two displacements), then **R** is the total displacement. The person taking the walk ends up at the tip of **R** . There are many ways to arrive at the same point. The person could have walked straight ahead first in the x-direction and then in the y-direction. Those paths are the x- and y-components of the resultant, $\mathbf{R}_x$ and $\mathbf{R}_y$. If we know $\mathbf{R}_x$ and $\mathbf{R}_y$, we can find R and θ using the equations $R = \sqrt{R_x{}^2 + R_y{}^2}$ and $\theta = tan^{-1}(R_y/R_x)$.

1. Draw in the x and y components of each vector (including the resultant) with a dashed line. Use the equations $A_x = A\cos\theta$ and $A_y = A\sin\theta$ to find the components. In Figure 5.23, these components are A_x, A_y, B_x, and B_y. Vector **A** makes an angle of θ_A with the x-axis, and vector **B** makes and angle of θ_B with its own x-axis (which is slightly above the x-axis used by vector **A**).

Figure 5.23 To add vectors **A** and **B**, first determine the horizontal and vertical components of each vector. These are the dotted vectors $\mathbf{A}_x$, $\mathbf{A}_y$ $\mathbf{B}_y$ shown in the image.

2. Find the x component of the resultant by adding the x component of the vectors

$$R_x = A_x + B_x$$

and find the y component of the resultant (as illustrated in Figure 5.24) by adding the y component of the vectors.

$$R_y = A_y + B_y.$$

Figure 5.24 The vectors $\mathbf{A}_x$ and $\mathbf{B}_x$ add to give the magnitude of the resultant vector in the horizontal direction, R_x. Similarly, the vectors $\mathbf{A}_y$ and $\mathbf{B}_y$ add to give the magnitude of the resultant vector in the vertical direction, R_y.

Now that we know the components of $\mathbf{R}$, we can find its magnitude and direction.

3. To get the magnitude of the resultant R, use the Pythagorean theorem.

$$R = \sqrt{R_x^2 + R_y^2}$$

4. To get the direction of the resultant

$$\theta = \tan^{-1}(R_y/R_x).$$

WATCH PHYSICS

Classifying Vectors and Quantities Example

This video contrasts and compares three vectors in terms of their magnitudes, positions, and directions.

Click to view content (https://www.youtube.com/embed/YpoEhcVBxNU)

GRASP CHECK

Three vectors, $\vec{u}$, $\vec{v}$, and $\vec{w}$, have the same magnitude of 5 units. Vector $\vec{v}$ points to the northeast. Vector $\vec{w}$ points to the southwest exactly opposite to vector $\vec{u}$. Vector $\vec{u}$ points in the northwest. If the vectors $\vec{u}$, $\vec{v}$, and $\vec{w}$ were added together, what would be the magnitude of the resultant vector? Why?

a. 0 units. All of them will cancel each other out.
b. 5 units. Two of them will cancel each other out.
c. 10 units. Two of them will add together to give the resultant.
d. 15 units. All of them will add together to give the resultant.

TIPS FOR SUCCESS

In the video, the vectors were represented with an arrow above them rather than in bold. This is a common notation in math classes.

Using the Analytical Method of Vector Addition and Subtraction to Solve Problems

Figure 5.25 uses the analytical method to add vectors.

WORKED EXAMPLE

An Accelerating Subway Train

Add the vector **A** to the vector **B** shown in Figure 5.25, using the steps above. The *x*-axis is along the east–west direction, and the *y*-axis is along the north–south directions. A person first walks 53.0 m in a direction 20.0° north of east, represented by vector **A**. The person then walks 34.0 m in a direction 63.0° north of east, represented by vector **B**.

Figure 5.25 You can use analytical models to add vectors.

Strategy

The components of **A** and **B** along the *x*- and *y*-axes represent walking due east and due north to get to the same ending point. We will solve for these components and then add them in the x-direction and y-direction to find the resultant.

Solution

First, we find the components of **A** and **B** along the *x*- and *y*-axes. From the problem, we know that $A = 53.0\ \text{m}$, $\theta_A = 20.0°$, $B = 34.0\ \text{m}$, and $\theta_B = 63.0°$. We find the *x*-components by using $A_x = A\cos\theta$, which gives

$$\begin{aligned} A_x &= A\cos\theta_A = (53.0\ \text{m})(\cos 20.0°) \\ &= (53.0\ \text{m})(0.940) = 49.8\ \text{m} \end{aligned}$$

and

$$\begin{aligned} B_x &= B\cos\theta_B = (34.0\ \text{m})(\cos 63.0°) \\ &= (34.0\ \text{m})(0.454) = 15.4\ \text{m}. \end{aligned}$$

Similarly, the *y*-components are found using $A_y = A\sin\theta_A$

$$\begin{aligned} A_y &= A\sin\theta_A = (53.0\ \text{m})(\sin 20.0°) \\ &= (53.0\ \text{m})(0.342) = 18.1\ \text{m} \end{aligned}$$

and

$$\begin{aligned} B_y &= B\sin\theta_B = (34.0\ \text{m})(\sin 63.0°) \\ &= (34.0\ \text{m})(0.891) = 30.3\ \text{m}. \end{aligned}$$

The *x*- and *y*-components of the resultant are

$$R_x = A_x + B_x = 49.8\ \text{m} + 15.4\ \text{m} = 65.2\ \text{m}$$

and

$$R_y = A_y + B_y = 18.1 \text{ m} + 30.3 \text{ m} = 48.4 \text{ m}.$$

Now we can find the magnitude of the resultant by using the Pythagorean theorem

$$R = \sqrt{R_x^2 + R_y^2} = \sqrt{(65.2)^2 + (48.4)^2} \text{ m} \tag{5.8}$$

so that

$$R = \sqrt{6601 \text{ m}} = 81.2 \text{ m}.$$

Finally, we find the direction of the resultant

$$\theta = \tan^{-1}(R_y/R_x) = +\tan^{-1}(48.4/65.2).$$

This is

$$\theta = \tan^{-1}(0.742) = 36.6^\circ.$$

Discussion

This example shows vector addition using the analytical method. Vector subtraction using the analytical method is very similar. It is just the addition of a negative vector. That is, $\mathrm{A} - \mathrm{B} \equiv \mathrm{A} + (-\mathrm{B})$. The components of $-\mathrm{B}$ are the negatives of the components of B. Therefore, the *x*- and *y*-components of the resultant $\mathrm{A} - \mathrm{B} = \mathrm{R}$ are

$$R_x = A_x + -B_x$$

and

$$R_y = A_y + -B_y$$

and the rest of the method outlined above is identical to that for addition.

Practice Problems

5. What is the magnitude of a vector whose *x*-component is 4 cm and whose *y*-component is 3 cm?
   a. 1 cm
   b. 5 cm
   c. 7 cm
   d. 25 cm

6. What is the magnitude of a vector that makes an angle of 30° to the horizontal and whose *x*-component is 3 units?
   a. 2.61 units
   b. 3.00 units
   c. 3.46 units
   d. 6.00 units

LINKS TO PHYSICS

Atmospheric Science

Figure 5.26 This picture shows Bert Foord during a television Weather Forecast from the Meteorological Office in 1963. (BBC TV)

Atmospheric science is a physical science, meaning that it is a science based heavily on physics. Atmospheric science includes meteorology (the study of weather) and climatology (the study of climate). Climate is basically the average weather over a longer time scale. Weather changes quickly over time, whereas the climate changes more gradually.

The movement of air, water and heat is vitally important to climatology and meteorology. Since motion is such a major factor in weather and climate, this field uses vectors for much of its math.

Vectors are used to represent currents in the ocean, wind velocity and forces acting on a parcel of air. You have probably seen a weather map using vectors to show the strength (magnitude) and direction of the wind.

Vectors used in atmospheric science are often three-dimensional. We won't cover three-dimensional motion in this text, but to go from two-dimensions to three-dimensions, you simply add a third vector component. Three-dimensional motion is represented as a combination of x-, y- and z components, where z is the altitude.

Vector calculus combines vector math with calculus, and is often used to find the rates of change in temperature, pressure or wind speed over time or distance. This is useful information, since atmospheric motion is driven by changes in pressure or temperature. The greater the variation in pressure over a given distance, the stronger the wind to try to correct that imbalance. Cold air tends to be more dense and therefore has higher pressure than warm air. Higher pressure air rushes into a region of lower pressure and gets deflected by the spinning of the Earth, and friction slows the wind at Earth's surface.

Finding how wind changes over distance and multiplying vectors lets meteorologists, like the one shown in Figure 5.26, figure out how much rotation (spin) there is in the atmosphere at any given time and location. This is an important tool for tornado prediction. Conditions with greater rotation are more likely to produce tornadoes.

GRASP CHECK

Why are vectors used so frequently in atmospheric science?

a. Vectors have magnitude as well as direction and can be quickly solved through scalar algebraic operations.
b. Vectors have magnitude but no direction, so it becomes easy to express physical quantities involved in the atmospheric science.
c. Vectors can be solved very accurately through geometry, which helps to make better predictions in atmospheric science.
d. Vectors have magnitude as well as direction and are used in equations that describe the three dimensional motion of the atmosphere.

Check Your Understanding

7. Between the analytical and graphical methods of vector additions, which is more accurate? Why?
   a. The analytical method is less accurate than the graphical method, because the former involves geometry and

trigonometry.

b. The analytical method is more accurate than the graphical method, because the latter involves some extensive calculations.

c. The analytical method is less accurate than the graphical method, because the former includes drawing all figures to the right scale.

d. The analytical method is more accurate than the graphical method, because the latter is limited by the precision of the drawing.

8. What is a component of a two dimensional vector?

a. A component is a piece of a vector that points in either the *x* or *y* direction.

b. A component is a piece of a vector that has half of the magnitude of the original vector.

c. A component is a piece of a vector that points in the direction opposite to the original vector.

d. A component is a piece of a vector that points in the same direction as original vector but with double of its magnitude.

9. How can we determine the global angle θ (measured counter-clockwise from positive *x*) if we know A_x and A_y?

a. $\theta = \cos^{-1} \frac{A_y}{A_x}$

b. $\theta = \cot^{-1} \frac{A_y}{A_x}$

c. $\theta = \sin^{-1} \frac{A_y}{A_x}$

d. $\theta = \tan^{-1} \frac{A_y}{A_x}$

10. How can we determine the magnitude of a vector if we know the magnitudes of its components?

a. $\left|\vec{A}\right| = A_x + A_y$

b. $\left|\vec{A}\right| = {A_x}^2 + {A_y}^2$

c. $\left|\vec{A}\right| = \sqrt{{A_x}^2 + {A_y}^2}$

d. $\left|\vec{A}\right| = \left({A_x}^2 + {A_y}^2\right)^2$

5.3 Projectile Motion

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Describe the properties of projectile motion
- Apply kinematic equations and vectors to solve problems involving projectile motion

Section Key Terms

air resistance	maximum height (of a projectile)	projectile
projectile motion	range	trajectory

Properties of Projectile Motion

Projectile motion is the motion of an object thrown (projected) into the air. After the initial force that launches the object, it only experiences the force of gravity. The object is called a **projectile**, and its path is called its **trajectory**. As an object travels through the air, it encounters a frictional force that slows its motion called **air resistance**. Air resistance does significantly alter trajectory motion, but due to the difficulty in calculation, it is ignored in introductory physics.

The most important concept in projectile motion is that *horizontal and vertical motions are independent*, meaning that they don't influence one another. Figure 5.27 compares a cannonball in free fall (in blue) to a cannonball launched horizontally in projectile motion (in red). You can see that the cannonball in free fall falls at the same rate as the cannonball in projectile motion. Keep in mind that if the cannon launched the ball with any vertical component to the velocity, the vertical displacements would not line up perfectly.

Since vertical and horizontal motions are independent, we can analyze them separately, along perpendicular axes. To do this, we separate projectile motion into the two components of its motion, one along the horizontal axis and the other along the vertical.

Figure 5.27 The diagram shows the projectile motion of a cannonball shot at a horizontal angle versus one dropped with no horizontal velocity. Note that both cannonballs have the same vertical position over time.

We'll call the horizontal axis the *x*-axis and the vertical axis the *y*-axis. For notation, **d** is the total displacement, and **x** and **y** are its components along the horizontal and vertical axes. The magnitudes of these vectors are *x* and *y*, as illustrated in Figure 5.28.

Figure 5.28 A boy kicks a ball at angle θ, and it is displaced a distance of **s** along its trajectory.

As usual, we use velocity, acceleration, and displacement to describe motion. We must also find the components of these variables along the *x*- and *y*-axes. The components of acceleration are then very simple $\mathbf{a_y} = -\mathbf{g} = -9.80\ \text{m/s}^2$. Note that this definition defines the upwards direction as positive. Because gravity is vertical, $\mathbf{a_x} = 0$. Both accelerations are constant, so we can use the kinematic equations. For review, the kinematic equations from a previous chapter are summarized in Table 5.1.

$\mathbf{x} = \mathbf{x}_0 + \mathbf{v}_{\mathbf{avg}} t$ (when $\mathbf{a}=0$)
$\mathbf{v}_{\mathbf{avg}} = \frac{\mathbf{v}_0+\mathbf{v}}{2}$ (when $\mathbf{a}=0$)
$\mathbf{v} = \mathbf{v}_0 + \mathbf{a}t$
$\mathbf{x} = \mathbf{x}_0 + \mathbf{v}_0 t + \frac{1}{2}\mathbf{a}t^2$
$\mathbf{v}^2 = \mathbf{v}_0^2 + 2\mathbf{a}(\mathbf{x} - \mathbf{x}_0)$

Table 5.1 Summary of Kinematic Equations (constant a)

Where **x** is position, $\mathbf{x}_o$ is initial position, **v** is velocity, $\mathbf{v}_{avg}$ is average velocity, *t* is time and a is acceleration.

Solve Problems Involving Projectile Motion

The following steps are used to analyze projectile motion:

1. Separate the motion into horizontal and vertical components along the x- and y-axes. These axes are perpendicular, so $A_x = A\cos\theta$ and $A_y = A\sin\theta$ are used. The magnitudes of the displacement $\mathbf{s}$ along x- and y-axes are called x and y. The magnitudes of the components of the velocity $\mathbf{v}$ are $v_x = v \quad \cos\theta$ and $v_y = v \quad \sin\theta$, where v is the magnitude of the velocity and θ is its direction. Initial values are denoted with a subscript 0.
2. Treat the motion as two independent one-dimensional motions, one horizontal and the other vertical. The kinematic equations for horizontal and vertical motion take the following forms

$$\text{Horizontal Motion}(\mathbf{a}_x = 0)$$
$$x = x_0 + v_x t$$
$$v_x = v_{0x} = \mathbf{v}_\text{x} = \text{velocity is a constant.}$$

Vertical motion (assuming positive is up $\mathbf{a}_y = -\mathbf{g} = -9.80\ \text{m/s}^2$)

$$\begin{aligned} y &= y_0 + \tfrac{1}{2}(v_{0y} + v_y)t \\ v_y &= v_{0y} - \mathbf{g}t \\ y &= y_0 + v_{0y}t - \tfrac{1}{2}\mathbf{g}t^2 \\ v_y^2 &= v_{0y}^2 - 2g(y - y_0) \end{aligned}$$

3. Solve for the unknowns in the two separate motions (one horizontal and one vertical). Note that the only common variable between the motions is time t . The problem solving procedures here are the same as for one-dimensional kinematics.
4. Recombine the two motions to find the total displacement $\mathbf{s}$ and velocity $\mathbf{v}$. We can use the analytical method of vector addition, which uses $A = \sqrt{{A_x}^2 + {A_y}^2}$ and $\theta = \tan^{-1}(A_y/A_x)$ to find the magnitude and direction of the total displacement and velocity.

$$\text{Displacement}$$
$$\mathbf{d} = \sqrt{x^2 + y^2}$$
$$\theta = \tan^{-1}(y/x)$$
$$\text{Velocity}$$
$$\mathbf{v} = \sqrt{\mathbf{v}_x^2 + \mathbf{v}_y^2}$$
$$\theta_v = \tan^{-1}(\mathbf{v}_y/\mathbf{v}_x)$$

θ is the direction of the displacement $\mathbf{d}$, and θ_v is the direction of the velocity $\mathbf{v}$. (See Figure 5.29

Figure 5.29 (a) We analyze two-dimensional projectile motion by breaking it into two independent one-dimensional motions along the vertical and horizontal axes. (b) The horizontal motion is simple, because $\mathbf{a}_x = 0$ and v_x is thus constant. (c) The velocity in the vertical direction begins to decrease as the object rises; at its highest point, the vertical velocity is zero. As the object falls towards the Earth again, the vertical velocity increases again in magnitude but points in the opposite direction to the initial vertical velocity. (d) The *x*- and *y*-motions are recombined to give the total velocity at any given point on the trajectory.

TIPS FOR SUCCESS

For problems of projectile motion, it is important to set up a coordinate system. The first step is to choose an initial position for $\mathbf{x}$ and $\mathbf{y}$. Usually, it is simplest to set the initial position of the object so that $\mathbf{x}_0 = 0$ and $\mathbf{y}_0 = 0$.

WATCH PHYSICS

Projectile at an Angle

This video presents an example of finding the displacement (or range) of a projectile launched at an angle. It also reviews basic trigonometry for finding the sine, cosine and tangent of an angle.

Click to view content (https://www.khanacademy.org/embed_video?v=ZZ3901rAZWY)

GRASP CHECK

Assume the ground is uniformly level. If the horizontal component a projectile's velocity is doubled, but the vertical component is unchanged, what is the effect on the time of flight?

a. The time to reach the ground would remain the same since the vertical component is unchanged.
b. The time to reach the ground would remain the same since the vertical component of the velocity also gets doubled.
c. The time to reach the ground would be halved since the horizontal component of the velocity is doubled.
d. The time to reach the ground would be doubled since the horizontal component of the velocity is doubled.

WORKED EXAMPLE

A Fireworks Projectile Explodes High and Away

During a fireworks display like the one illustrated in Figure 5.30, a shell is shot into the air with an initial speed of 70.0 m/s at an angle of 75° above the horizontal. The fuse is timed to ignite the shell just as it reaches its highest point above the ground. (a) Calculate the height at which the shell explodes. (b) How much time passed between the launch of the shell and the explosion? (c) What is the horizontal displacement of the shell when it explodes?

Figure 5.30 The diagram shows the trajectory of a fireworks shell.

Strategy

The motion can be broken into horizontal and vertical motions in which $\mathbf{a}_x = 0$ and $\mathbf{a}_y = \mathbf{g}$. We can then define $\mathbf{x}_0$ and $\mathbf{y}_0$ to be zero and solve for the **maximum height**.

Solution for (a)

By height we mean the altitude or vertical position $\mathbf{y}$ above the starting point. The highest point in any trajectory, the maximum height, is reached when $\mathbf{v}_y = 0$; this is the moment when the vertical velocity switches from positive (upwards) to negative (downwards). Since we know the initial velocity, initial position, and the value of $\mathbf{v}_y$ when the firework reaches its maximum height, we use the following equation to find $\mathbf{y}$

$$\mathbf{v}_y^2 = \mathbf{v}_{0y}^2 - 2\mathbf{g}(\mathbf{y} - y_0).$$

Because $\mathbf{y}_0$ and $\mathbf{v}_y$ are both zero, the equation simplifies to

$$0 = \mathbf{v}_{0y}^2 - 2\mathbf{g}\mathbf{y}.$$

Solving for $\mathbf{y}$ gives

$$\mathbf{y} = \frac{\mathbf{v}_{0y}^2}{2\mathbf{g}}.$$

Now we must find $\mathbf{v}_{0y}$, the component of the initial velocity in the y-direction. It is given by $\mathbf{v}_{0y} = \mathbf{v}_0 \sin\theta$, where $\mathbf{v}_{0y}$ is the initial velocity of 70.0 m/s, and $\theta = 75^\circ$ is the initial angle. Thus,

$$\mathbf{v}_{0y} = \mathbf{v}_0 \sin\theta_0 = (70.0 \text{ m/s})(\sin 75^\circ) = 67.6 \text{ m/s}$$

and $\mathbf{y}$ is

$$\mathbf{y} = \frac{(67.6 \text{ m/s})^2}{2(9.80 \text{ m/s}^2),}$$

so that

$$\mathbf{y} = 233 \text{ m}.$$

Discussion for (a)

Since up is positive, the initial velocity and maximum height are positive, but the acceleration due to gravity is negative. The maximum height depends only on the vertical component of the initial velocity. The numbers in this example are reasonable for large fireworks displays, the shells of which do reach such heights before exploding.

Solution for (b)

There is more than one way to solve for the time to the highest point. In this case, the easiest method is to use $\mathbf{y} = \mathbf{y}_0 + \frac{1}{2}(\mathbf{v}_{0y} + \mathbf{v}_y)t$. Because y_0 is zero, this equation reduces to

$$\mathbf{y} = \frac{1}{2}(\mathbf{v}_{0y} + \mathbf{v}_y)t.$$

Note that the final vertical velocity, $\mathbf{v}_y$, at the highest point is zero. Therefore,

$$\begin{aligned} t &= \frac{2\mathbf{y}}{(\mathbf{v}_{0\mathbf{y}} + \mathbf{v}_y)} = \frac{2(233 \text{ m})}{(67.6 \text{ m/s})} \\ &= 6.90 \text{ s}. \end{aligned}$$

Discussion for (b)

This time is also reasonable for large fireworks. When you are able to see the launch of fireworks, you will notice several seconds pass before the shell explodes. Another way of finding the time is by using $\mathbf{y} = \mathbf{y}_0 + \mathbf{v}_{0\mathbf{y}}t - \frac{1}{2}\mathbf{g}t^2$, and solving the quadratic equation for t .

Solution for (c)

Because air resistance is negligible, $\mathbf{a}_x = 0$ and the horizontal velocity is constant. The horizontal displacement is horizontal velocity multiplied by time as given by $\mathbf{x} = \mathbf{x}_0 + \mathbf{v}_x t$, where $\mathbf{x}_0$ is equal to zero

$$\mathbf{x} = \mathbf{v}_x t,$$

where $\mathbf{v}_x$ is the x-component of the velocity, which is given by $\mathbf{v}_x = \mathbf{v}_0 \cos\theta_0$. Now,

$$\mathbf{v}_x = \mathbf{v}_0 \cos\theta_0 = (70.0 \text{ m/s})(\cos 75^\circ) = 18.1 \text{ m/s}.$$

The time t for both motions is the same, and so $\mathbf{x}$ is

$$\mathbf{x} = (18.1 \text{ m/s})(6.90 \text{ s}) = 125 \text{ m}.$$

Discussion for (c)

The horizontal motion is a constant velocity in the absence of air resistance. The horizontal displacement found here could be useful in keeping the fireworks fragments from falling on spectators. Once the shell explodes, air resistance has a major effect, and many fragments will land directly below, while some of the fragments may now have a velocity in the –x direction due to the

forces of the explosion.

The expression we found for $\mathbf{y}$ while solving part (a) of the previous problem works for any projectile motion problem where air resistance is negligible. Call the maximum height $\mathbf{y} = h$; then,

$$h = \frac{\mathbf{v}_{0y}^2}{2\mathbf{g}}.$$

This equation defines the **maximum height of a projectile**. The maximum height depends only on the vertical component of the initial velocity.

WORKED EXAMPLE

Calculating Projectile Motion: Hot Rock Projectile

Suppose a large rock is ejected from a volcano, as illustrated in Figure 5.31, with a speed of 25.0 m/s and at an angle 35° above the horizontal. The rock strikes the side of the volcano at an altitude 20.0 m lower than its starting point. (a) Calculate the time it takes the rock to follow this path.

Figure 5.31 The diagram shows the projectile motion of a large rock from a volcano.

Strategy

Breaking this two-dimensional motion into two independent one-dimensional motions will allow us to solve for the time. The time a projectile is in the air depends only on its vertical motion.

Solution

While the rock is in the air, it rises and then falls to a final position 20.0 m lower than its starting altitude. We can find the time for this by using

$$\mathbf{y} = \mathbf{y}_0 + \mathbf{v}_{0\mathbf{y}} t - \frac{1}{2}\mathbf{g}t^2.$$

If we take the initial position $\mathbf{y}_0$ to be zero, then the final position is $\mathbf{y} = -20.0\ \text{m}$. Now the initial vertical velocity is the vertical component of the initial velocity, found from

$$\mathbf{v}_{0y} = \mathbf{v}_0 \sin\theta_0 = (25.0\ \text{m/s})(\sin 35^\circ) = 14.3\ \text{m/s}. \tag{5.9}$$

Substituting known values yields

$$-20.0\ \text{m} = (14.3\ \text{m/s})t - \left(4.90\ \text{m/s}^2\right) t^2.$$

Rearranging terms gives a quadratic equation in t

$$\left(4.90\ \text{m/s}^2\right) t^2 - (14.3\ \text{m/s})\, t - (20.0\ \text{m}) = 0.$$

This expression is a quadratic equation of the form $at^2 + bt + c = 0$, where the constants are $a = 4.90$, $b = -14.3$, and $c = -20.0$. Its solutions are given by the quadratic formula

$$t = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

This equation yields two solutions $t = 3.96$ and $t = -1.03$. You may verify these solutions as an exercise. The time is t = 3.96 s or −1.03 s. The negative value of time implies an event before the start of motion, so we discard it. Therefore,

$$t = 3.96 \text{ s}.$$

Discussion

The time for projectile motion is completely determined by the vertical motion. So any projectile that has an initial vertical velocity of 14.3m/s and lands 20.0 m below its starting altitude will spend 3.96 s in the air.

Practice Problems

11. If an object is thrown horizontally, travels with an average x-component of its velocity equal to 5 m/s, and does not hit the ground, what will be the x-component of the displacement after 20 s?

a. -100 m
b. -4 m
c. 4 m
d. 100 m

12. If a ball is thrown straight up with an initial velocity of 20 m/s upward, what is the maximum height it will reach?

a. -20.4 m
b. -1.02 m
c. 1.02 m
d. 20.4 m

The fact that vertical and horizontal motions are independent of each other lets us predict the range of a projectile. The **range** is the horizontal distance **R** traveled by a projectile on level ground, as illustrated in Figure 5.32. Throughout history, people have been interested in finding the range of projectiles for practical purposes, such as aiming cannons.

Figure 5.32 Trajectories of projectiles on level ground. (a) The greater the initial speed v_0, the greater the range for a given initial angle. (b) The effect of initial angle θ_0 on the range of a projectile with a given initial speed. Note that any combination of trajectories that add to 90 degrees will have the same range in the absence of air resistance, although the maximum heights of those paths are different.

How does the initial velocity of a projectile affect its range? Obviously, the greater the initial speed v_0, the greater the range, as shown in the figure above. The initial angle θ_0 also has a dramatic effect on the range. When air resistance is negligible, the range R of a projectile on *level ground* is

$$R = \frac{v_0^2 \sin 2\theta_0}{\mathbf{g}},$$

where v_0 is the initial speed and θ_0 is the initial angle relative to the horizontal. It is important to note that the range doesn't apply to problems where the initial and final y position are different, or to cases where the object is launched perfectly horizontally.

Virtual Physics

Projectile Motion

In this simulation you will learn about projectile motion by blasting objects out of a cannon. You can choose between objects such as a tank shell, a golf ball or even a Buick. Experiment with changing the angle, initial speed, and mass, and adding in air resistance. Make a game out of this simulation by trying to hit the target.

Click to view content (https://archive.cnx.org/specials/317dbd00-8e61-4065-b3eb-f2b80db9b7ed/projectile-motion/)

GRASP CHECK

If a projectile is launched on level ground, what launch angle maximizes the range of the projectile?

a. 0°
b. 30°
c. 45°
d. 60°

Check Your Understanding

13. What is projectile motion?

a. Projectile motion is the motion of an object projected into the air, which moves under the influence of gravity.
b. Projectile motion is the motion of an object projected into the air which moves independently of gravity.
c. Projectile motion is the motion of an object projected vertically upward into the air which moves under the influence of gravity.
d. Projectile motion is the motion of an object projected horizontally into the air which moves independently of gravity.

14. What is the force experienced by a projectile after the initial force that launched it into the air in the absence of air resistance?

a. The nuclear force
b. The gravitational force
c. The electromagnetic force
d. The contact force

5.4 Inclined Planes

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Distinguish between static friction and kinetic friction
- Solve problems involving inclined planes

Section Key Terms

kinetic friction	static friction

Static Friction and Kinetic Friction

Recall from the previous chapter that friction is a force that opposes motion, and is around us all the time. Friction allows us to move, which you have discovered if you have ever tried to walk on ice.

There are different types of friction—kinetic and static. **Kinetic friction** acts on an object in motion, while **static friction** acts on an object or system at rest. The maximum static friction is usually greater than the kinetic friction between the objects.

Imagine, for example, trying to slide a heavy crate across a concrete floor. You may push harder and harder on the crate and not move it at all. This means that the static friction responds to what you do—it increases to be equal to and in the opposite direction of your push. But if you finally push hard enough, the crate seems to slip suddenly and starts to move. Once in motion, it is easier to keep it in motion than it was to get it started because the kinetic friction force is less than the static friction force. If you were to add mass to the crate, (for example, by placing a box on top of it) you would need to push even harder to get it started and also to keep it moving. If, on the other hand, you oiled the concrete you would find it easier to get the crate started and keep it going.

Figure 5.33 shows how friction occurs at the interface between two objects. Magnifying these surfaces shows that they are rough on the microscopic level. So when you push to get an object moving (in this case, a crate), you must raise the object until it can skip along with just the tips of the surface hitting, break off the points, or do both. The harder the surfaces are pushed together (such as if another box is placed on the crate), the more force is needed to move them.

Figure 5.33 Frictional forces, such as **f**, always oppose motion or attempted motion between objects in contact. Friction arises in part because of the roughness of the surfaces in contact, as seen in the expanded view.

The magnitude of the frictional force has two forms: one for static friction, the other for kinetic friction. When there is no motion between the objects, the magnitude of static friction $\mathbf{f_s}$ is

$$\mathbf{f}_s \leq \mu_s \mathbf{N}_s,$$

where μ_s is the coefficient of static friction and **N** is the magnitude of the normal force. Recall that the normal force opposes the force of gravity and acts perpendicular to the surface in this example, but not always.

Since the symbol $\leq$ means less than or equal to, this equation says that static friction can have a maximum value of $\mu_s \mathbf{N}$. That is,

$$\mathbf{f}_s(\text{max}) = \mu_s \mathbf{N}.$$

Static friction is a responsive force that increases to be equal and opposite to whatever force is exerted, up to its maximum limit. Once the applied force exceeds **f**s(max), the object will move. Once an object is moving, the magnitude of kinetic friction $\mathbf{f}_k$ is given by

$$\mathbf{f}_k = \mu_k \mathbf{N}.$$

where μ_k is the coefficient of kinetic friction.

Friction varies from surface to surface because different substances are rougher than others. Table 5.2 compares values of static and kinetic friction for different surfaces. The coefficient of the friction depends on the two surfaces that are in contact.

System	Static Friction μ_s	Kinetic Friction μ_k
Rubber on dry concrete	1.0	0.7
Rubber on wet concrete	0.7	0.5
Wood on wood	0.5	0.3
Waxed wood on wet snow	0.14	0.1
Metal on wood	0.5	0.3
Steel on steel (dry)	0.6	0.3
Steel on steel (oiled)	0.05	0.03
Teflon on steel	0.04	0.04
Bone lubricated by synovial fluid	0.016	0.015
Shoes on wood	0.9	0.7

Table 5.2 Coefficients of Static and Kinetic Friction

System	Static Friction μ_s	Kinetic Friction μ_k
Shoes on ice	0.1	0.05
Ice on ice	0.1	0.03
Steel on ice	0.4	0.02

Table 5.2 Coefficients of Static and Kinetic Friction

Since the direction of friction is always opposite to the direction of motion, friction runs parallel to the surface between objects and perpendicular to the normal force. For example, if the crate you try to push (with a force parallel to the floor) has a mass of 100 kg, then the normal force would be equal to its weight

$$\mathbf{W} = m\mathbf{g} = (100\text{ kg})(9.80\text{m/s}^2) = 980\ \text{N},$$

perpendicular to the floor. If the coefficient of static friction is 0.45, you would have to exert a force parallel to the floor greater than

$$\mathbf{f}_s(\text{max}) = \mu_s\mathbf{N} = (0.45)(980\ \text{N}) = 440\ \text{N}$$

to move the crate. Once there is motion, friction is less and the coefficient of kinetic friction might be 0.30, so that a force of only 290 N

$$\mathbf{f}_k = \mu_k\mathbf{N} = (0.30)(980\ \text{N}) = 290\ \text{N}$$

would keep it moving at a constant speed. If the floor were lubricated, both coefficients would be much smaller than they would be without lubrication. The coefficient of friction is unitless and is a number usually between 0 and 1.0.

Working with Inclined Planes

We discussed previously that when an object rests on a horizontal surface, there is a normal force supporting it equal in magnitude to its weight. Up until now, we dealt only with normal force in one dimension, with gravity and normal force acting perpendicular to the surface in opposing directions (gravity downward, and normal force upward). Now that you have the skills to work with forces in two dimensions, we can explore what happens to weight and the normal force on a tilted surface such as an inclined plane. For inclined plane problems, it is easier breaking down the forces into their components if we rotate the coordinate system, as illustrated in Figure 5.34. The first step when setting up the problem is to break down the force of weight into components.

Figure 5.34 The diagram shows perpendicular and horizontal components of weight on an inclined plane.

When an object rests on an incline that makes an angle θ with the horizontal, the force of gravity acting on the object is divided into two components: A force acting perpendicular to the plane, $\mathbf{w}_\perp$, and a force acting parallel to the plane, $\mathbf{w}_{||}$. The perpendicular force of weight, $\mathbf{w}_\perp$, is typically equal in magnitude and opposite in direction to the normal force, $\mathbf{N}$. The force acting parallel to the plane, $\mathbf{w}_{||}$, causes the object to accelerate down the incline. The force of friction, $\mathbf{f}$, opposes the motion of the object, so it acts upward along the plane.

It is important to be careful when resolving the weight of the object into components. If the angle of the incline is at an angle θ to the horizontal, then the magnitudes of the weight components are

$$\mathbf{w}_{||} = \mathbf{w}sin(\theta) = m\mathbf{g}sin(\theta) \text{ and}$$
$$\mathbf{w}_{\perp} = \mathbf{w}cos(\theta) = m\mathbf{g}cos(\theta).$$

Instead of memorizing these equations, it is helpful to be able to determine them from reason. To do this, draw the right triangle formed by the three weight vectors. Notice that the angle of the incline is the same as the angle formed between $\mathbf{w}$ and $\mathbf{w}_{\perp}$. Knowing this property, you can use trigonometry to determine the magnitude of the weight components

$$cos(\theta) = \frac{\mathbf{w}_{\perp}}{\mathbf{w}}$$
$$\mathbf{w}_{\perp} = \mathbf{w}cos(\theta) = m\mathbf{g}cos(\theta)$$
$$sin(\theta) = \frac{\mathbf{w}_{||}}{\mathbf{w}}$$
$$\mathbf{w}_{||} = \mathbf{w}sin(\theta) = m\mathbf{g}sin(\theta).$$

WATCH PHYSICS

Inclined Plane Force Components

This video (https://www.khanacademy.org/embed_video?v=TC23wD34C7k) shows how the weight of an object on an inclined plane is broken down into components perpendicular and parallel to the surface of the plane. It explains the geometry for finding the angle in more detail.

GRASP CHECK

Click to view content (https://www.youtube.com/embed/TC23wD34C7k)

This video shows how the weight of an object on an inclined plane is broken down into components perpendicular and parallel to the surface of the plane. It explains the geometry for finding the angle in more detail.

When the surface is flat, you could say that one of the components of the gravitational force is zero; Which one? As the angle of the incline gets larger, what happens to the magnitudes of the perpendicular and parallel components of gravitational force?

a. When the angle is zero, the parallel component is zero and the perpendicular component is at a maximum. As the angle increases, the parallel component decreases and the perpendicular component increases. This is because the cosine of the angle shrinks while the sine of the angle increases.
b. When the angle is zero, the parallel component is zero and the perpendicular component is at a maximum. As the angle increases, the parallel component decreases and the perpendicular component increases. This is because the cosine of the angle increases while the sine of the angle shrinks.
c. When the angle is zero, the parallel component is zero and the perpendicular component is at a maximum. As the angle increases, the parallel component increases and the perpendicular component decreases. This is because the cosine of the angle shrinks while the sine of the angle increases.
d. When the angle is zero, the parallel component is zero and the perpendicular component is at a maximum. As the angle increases, the parallel component increases and the perpendicular component decreases. This is because the cosine of the angle increases while the sine of the angle shrinks.

TIPS FOR SUCCESS

Normal force is represented by the variable $\mathbf{N}$. This should not be confused with the symbol for the newton, which is also represented by the letter N. It is important to tell apart these symbols, especially since the units for normal force ($\mathbf{N}$) happen to be newtons (N). For example, the normal force, $\mathbf{N}$, that the floor exerts on a chair might be $\mathbf{N} = 100 \text{ N}$. One important difference is that normal force is a vector, while the newton is simply a unit. Be careful not to confuse these letters in your calculations!

To review, the process for solving inclined plane problems is as follows:

1. Draw a sketch of the problem.
2. Identify known and unknown quantities, and identify the system of interest.
3. Draw a free-body diagram (which is a sketch showing all of the forces acting on an object) with the coordinate system rotated at the same angle as the inclined plane. Resolve the vectors into horizontal and vertical components and draw them on the free-body diagram.
4. Write Newton's second law in the horizontal and vertical directions and add the forces acting on the object. If the object does not accelerate in a particular direction (for example, the x-direction) then $\mathbf{F}\text{net}\ x = 0$. If the object does accelerate in that direction, $\mathbf{F}\text{net}\ x = m\mathbf{a}$.
5. Check your answer. Is the answer reasonable? Are the units correct?

WORKED EXAMPLE

Finding the Coefficient of Kinetic Friction on an Inclined Plane

A skier, illustrated in Figure 5.35(a), with a mass of 62 kg is sliding down a snowy slope at an angle of 25 degrees. Find the coefficient of kinetic friction for the skier if friction is known to be 45.0 N.

Figure 5.35 Use the diagram to help find the coefficient of kinetic friction for the skier.

Strategy

The magnitude of kinetic friction was given as 45.0 N. Kinetic friction is related to the normal force $\mathbf{N}$ as $\mathbf{f}_k = \mu_k \mathbf{N}$. Therefore, we can find the coefficient of kinetic friction by first finding the normal force of the skier on a slope. The normal force is always perpendicular to the surface, and since there is no motion perpendicular to the surface, the normal force should equal the component of the skier's weight perpendicular to the slope.

That is,

$$\mathbf{N} = \mathbf{w}_\perp = \mathbf{w}\cos(25^\circ) = m\mathbf{g}\cos(25^\circ).$$

Substituting this into our expression for kinetic friction, we get

$$\mathbf{f}_k = \mu_k m\mathbf{g}\cos 25^\circ,$$

which can now be solved for the coefficient of kinetic friction μ_k.

Solution

Solving for μ_k gives

$$\mu_k = \frac{\mathbf{f}_k}{\mathbf{w}\cos 25^\circ} = \frac{\mathbf{f}_k}{m\mathbf{g}\cos 25^\circ}.$$

Substituting known values on the right-hand side of the equation,

$$\mu_k = \frac{45.0\ \text{N}}{(62\ \text{kg})(9.80\ \text{m/s}^2)(0.906)} = 0.082.$$

Discussion

This result is a little smaller than the coefficient listed in Table 5.2 for waxed wood on snow, but it is still reasonable since values

of the coefficients of friction can vary greatly. In situations like this, where an object of mass m slides down a slope that makes an angle θ with the horizontal, friction is given by $\mathbf{f}_k = \mu_k m\mathbf{g}\cos\theta$.

WORKED EXAMPLE

Weight on an Incline, a Two-Dimensional Problem

The skier's mass, including equipment, is 60.0 kg. (See Figure 5.36(b).) (a) What is her acceleration if friction is negligible? (b) What is her acceleration if the frictional force is 45.0 N?

Figure 5.36 Now use the diagram to help find the skier's acceleration if friction is negligible and if the frictional force is 45.0 N.

Strategy

The most convenient coordinate system for motion on an incline is one that has one coordinate parallel to the slope and one perpendicular to the slope. Remember that motions along perpendicular axes are independent. We use the symbol ⊥ to mean perpendicular, and || to mean parallel.

The only external forces acting on the system are the skier's weight, friction, and the normal force exerted by the ski slope, labeled $\mathbf{w}$, $\mathbf{f}$, and $\mathbf{N}$ in the free-body diagram. $\mathbf{N}$ is always perpendicular to the slope and $\mathbf{f}$ is parallel to it. But $\mathbf{w}$ is not in the direction of either axis, so we must break it down into components along the chosen axes. We define $\mathbf{w}_{||}$ to be the component of weight parallel to the slope and $\mathbf{w}_{\perp}$ the component of weight perpendicular to the slope. Once this is done, we can consider the two separate problems of forces parallel to the slope and forces perpendicular to the slope.

Solution

The magnitude of the component of the weight parallel to the slope is $\mathbf{w}_{||} = \mathbf{w}\sin(25°) = m\mathbf{g}\sin(25°)$, and the magnitude of the component of the weight perpendicular to the slope is $\mathbf{w}_{\perp} = \mathbf{w}\cos(25°) = m\mathbf{g}\cos(25°)$.

(a) Neglecting friction: Since the acceleration is parallel to the slope, we only need to consider forces parallel to the slope. Forces perpendicular to the slope add to zero, since there is no acceleration in that direction. The forces parallel to the slope are the amount of the skier's weight parallel to the slope $\mathbf{w}_{||}$ and friction $\mathbf{f}$. Assuming no friction, by Newton's second law the acceleration parallel to the slope is

$$\mathbf{a}_{||} = \frac{\mathbf{F}_{\text{net }||}}{m},$$

Where the net force parallel to the slope $\mathbf{F}_{\text{net }||} = \mathbf{w}_{||} = m\mathbf{g}\sin(25°)$, so that

$$\begin{aligned}\mathbf{a}_{||} &= \frac{\mathbf{F}_{\text{net }||}}{m} = \frac{m\mathbf{g}\sin(25°)}{m} = \mathbf{g}\sin(25°)\\ &= (9.80\text{ m/s}^2)(0.423) = 4.14\text{ m/s}^2\end{aligned}$$

is the acceleration.

(b) Including friction: Here we now have a given value for friction, and we know its direction is parallel to the slope and it opposes motion between surfaces in contact. So the net external force is now

$$\mathbf{F}_{\text{net}\,||} = \mathbf{w}_{||} - \mathbf{f},$$

and substituting this into Newton's second law, $a_{||} = \frac{\mathbf{F}_{\text{net}\,||}}{m}$ gives

$$\mathbf{a}_{||} = \frac{\mathbf{F}_{\text{net}\,||}}{m} = \frac{\mathbf{w}_{||} - \mathbf{f}}{m} = \frac{m\mathbf{g}\sin(25°) - \mathbf{f}}{m}.$$

We substitute known values to get

$$\mathbf{a}_{||} = \frac{(60.0\text{ kg})(9.80\text{ m/s}^2)(0.423) - 45.0\text{ N}}{60.0\text{ kg}},$$

or

$$\mathbf{a}_{||} = 3.39\text{ m/s}^2,$$

which is the acceleration parallel to the incline when there is 45 N opposing friction.

Discussion

Since friction always opposes motion between surfaces, the acceleration is smaller when there is friction than when there is not.

Practice Problems

15. When an object sits on an inclined plane that makes an angle θ with the horizontal, what is the expression for the component of the objects weight force that is parallel to the incline?

a. $w_{||} = w\cos\theta$
b. $w_{||} = w\sin\theta$
c. $w_{||} = w\sin\theta - \cos\theta$
d. $w_{||} = w\cos\theta - \sin\theta$

16. An object with a mass of 5 kg rests on a plane inclined 30° from horizontal. What is the component of the weight force that is parallel to the incline?

a. 4.33 N
b. 5.0 N
c. 24.5 N
d. 42.43 N

Snap Lab

Friction at an Angle: Sliding a Coin

An object will slide down an inclined plane at a constant velocity if the net force on the object is zero. We can use this fact to measure the coefficient of kinetic friction between two objects. As shown in the first Worked Example, the kinetic friction on a slope $\mathbf{f}_k = \mu_k m\mathbf{g}\cos\theta$, and the component of the weight down the slope is equal to $m\mathbf{g}\sin\theta$. These forces act in opposite directions, so when they have equal magnitude, the acceleration is zero. Writing these out

$$\begin{aligned} \mathbf{f}_k &= \mathbf{Fg}_x \\ \mu_k m\mathbf{g}\cos\theta &= m\mathbf{g}\sin\theta. \end{aligned}$$

Solving for μ_k, since $\tan\theta = \sin\theta/\cos\theta$ we find that

$$\mu_k = \frac{m\mathbf{g}\sin\theta}{m\mathbf{g}\cos\theta} = \tan\theta. \tag{5.10}$$

- 1 coin
- 1 book
- 1 protractor

1. Put a coin flat on a book and tilt it until the coin slides at a constant velocity down the book. You might need to tap the book lightly to get the coin to move.

2. Measure the angle of tilt relative to the horizontal and find μ_k .

GRASP CHECK

True or False—If only the angles of two vectors are known, we can find the angle of their resultant addition vector.

a. True
b. False

Check Your Understanding

17. What is friction?

a. Friction is an internal force that opposes the relative motion of an object.
b. Friction is an internal force that accelerates an object's relative motion.
c. Friction is an external force that opposes the relative motion of an object.
d. Friction is an external force that increases the velocity of the relative motion of an object.

18. What are the two varieties of friction? What does each one act upon?

a. Kinetic and static friction both act on an object in motion.
b. Kinetic friction acts on an object in motion, while static friction acts on an object at rest.
c. Kinetic friction acts on an object at rest, while static friction acts on an object in motion.
d. Kinetic and static friction both act on an object at rest.

19. Between static and kinetic friction between two surfaces, which has a greater value? Why?

a. The kinetic friction has a greater value because the friction between the two surfaces is more when the two surfaces are in relative motion.
b. The static friction has a greater value because the friction between the two surfaces is more when the two surfaces are in relative motion.
c. The kinetic friction has a greater value because the friction between the two surfaces is less when the two surfaces are in relative motion.
d. The static friction has a greater value because the friction between the two surfaces is less when the two surfaces are in relative motion.

5.5 Simple Harmonic Motion

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Describe Hooke's law and Simple Harmonic Motion
- Describe periodic motion, oscillations, amplitude, frequency, and period
- Solve problems in simple harmonic motion involving springs and pendulums

Section Key Terms

amplitude	deformation	equilibrium position	frequency
Hooke's law	oscillate	period	periodic motion
restoring force	simple harmonic motion	simple pendulum	

Hooke's Law and Simple Harmonic Motion

Imagine a car parked against a wall. If a bulldozer pushes the car into the wall, the car will not move but it will noticeably change shape. A change in shape due to the application of a force is a **deformation**. Even very small forces are known to cause some deformation. For small deformations, two important things can happen. First, unlike the car and bulldozer example, the object returns to its original shape when the force is removed. Second, the size of the deformation is proportional to the force. This

second property is known as **Hooke's law**. In equation form, Hooke's law is

$$\mathbf{F} = -\mathbf{kx},$$

where **x** is the amount of deformation (the change in length, for example) produced by the restoring force **F**, and **k** is a constant that depends on the shape and composition of the object. The restoring force is the force that brings the object back to its equilibrium position; the minus sign is there because the restoring force acts in the direction opposite to the displacement. Note that the restoring force is proportional to the deformation **x**. The deformation can also be thought of as a displacement from equilibrium. It is a change in position due to a force. In the absence of force, the object would rest at its equilibrium position. The force constant **k** is related to the stiffness of a system. The larger the force constant, the stiffer the system. A stiffer system is more difficult to deform and requires a greater restoring force. The units of **k** are newtons per meter (N/m). One of the most common uses of Hooke's law is solving problems involving springs and pendulums, which we will cover at the end of this section.

Oscillations and Periodic Motion

What do an ocean buoy, a child in a swing, a guitar, and the beating of hearts all have in common? They all **oscillate**. That is, they move back and forth between two points, like the ruler illustrated in Figure 5.37. All oscillations involve force. For example, you push a child in a swing to get the motion started.

Figure 5.37 A ruler is displaced from its equilibrium position.

Newton's first law implies that an object oscillating back and forth is experiencing forces. Without force, the object would move in a straight line at a constant speed rather than oscillate. Consider, for example, plucking a plastic ruler to the left as shown in Figure 5.38. The deformation of the ruler creates a force in the opposite direction, known as a **restoring force**. Once released, the restoring force causes the ruler to move back toward its stable equilibrium position, where the net force on it is zero. However, by the time the ruler gets there, it gains momentum and continues to move to the right, producing the opposite deformation. It is then forced to the left, back through equilibrium, and the process is repeated until it gradually loses all of its energy. The simplest oscillations occur when the restoring force is directly proportional to displacement. Recall that Hooke's law describes this situation with the equation $\mathbf{F} = -\mathbf{kx}$. Therefore, Hooke's law describes and applies to the simplest case of oscillation, known as **simple harmonic motion**.

Figure 5.38 (a) The plastic ruler has been released, and the restoring force is returning the ruler to its equilibrium position. (b) The net force is zero at the equilibrium position, but the ruler has momentum and continues to move to the right. (c) The restoring force is in the opposite direction. It stops the ruler and moves it back toward equilibrium again. (d) Now the ruler has momentum to the left. (e) In the absence of damping (caused by frictional forces), the ruler reaches its original position. From there, the motion will repeat itself.

When you pluck a guitar string, the resulting sound has a steady tone and lasts a long time. Each vibration of the string takes the same time as the previous one. **Periodic motion** is a motion that repeats itself at regular time intervals, such as with an object bobbing up and down on a spring or a pendulum swinging back and forth. The time to complete one oscillation (a complete cycle of motion) remains constant and is called the **period** T. Its units are usually seconds.

Frequency f is the number of oscillations per unit time. The SI unit for frequency is the hertz (Hz), defined as the number of oscillations per second. The relationship between frequency and period is

$$f = 1/T.$$

As you can see from the equation, frequency and period are different ways of expressing the same concept. For example, if you get a paycheck twice a month, you could say that the frequency of payment is two per month, or that the period between checks is half a month.

If there is no friction to slow it down, then an object in simple motion will oscillate forever with equal displacement on either side of the equilibrium position. The **equilibrium position** is where the object would naturally rest in the absence of force. The maximum displacement from equilibrium is called the **amplitude X**. The units for amplitude and displacement are the same, but depend on the type of oscillation. For the object on the spring, shown in Figure 5.39, the units of amplitude and displacement are meters.

Figure 5.39 An object attached to a spring sliding on a frictionless surface is a simple harmonic oscillator. When displaced from equilibrium, the object performs simple harmonic motion that has an amplitude **X** and a period T. The object's maximum speed occurs as it passes through equilibrium. The stiffer the spring is, the smaller the period T. The greater the mass of the object is, the greater the period T.

The mass m and the force constant **k** are the *only* factors that affect the period and frequency of simple harmonic motion. The period of a simple harmonic oscillator is given by

$$T = 2\pi\sqrt{\frac{m}{\mathbf{k}}}$$

and, because $f = 1/T$, the frequency of a simple harmonic oscillator is

$$f = \frac{1}{2\pi}\sqrt{\frac{\mathbf{k}}{m}}.$$

WATCH PHYSICS

Introduction to Harmonic Motion

This video shows how to graph the displacement of a spring in the x-direction over time, based on the period. Watch the first 10 minutes of the video (you can stop when the narrator begins to cover calculus).

Click to view content (https://www.khanacademy.org/embed_video?v=Nk2q-_jkJVs)

GRASP CHECK

If the amplitude of the displacement of a spring were larger, how would this affect the graph of displacement over time? What would happen to the graph if the period was longer?

a. Larger amplitude would result in taller peaks and troughs and a longer period would result in greater separation in time between peaks.
b. Larger amplitude would result in smaller peaks and troughs and a longer period would result in greater distance between peaks.
c. Larger amplitude would result in taller peaks and troughs and a longer period would result in shorter distance between peaks.
d. Larger amplitude would result in smaller peaks and troughs and a longer period would result in shorter distance between peaks.

Solving Spring and Pendulum Problems with Simple Harmonic Motion

Before solving problems with springs and pendulums, it is important to first get an understanding of how a pendulum works. Figure 5.40 provides a useful illustration of a simple pendulum.

Figure 5.40 A simple pendulum has a small-diameter bob and a string that has a very small mass but is strong enough not to stretch. The linear displacement from equilibrium is s, the length of the arc. Also shown are the forces on the bob, which result in a net force of $-mg\sin\theta$ toward the equilibrium position—that is, a restoring force.

Everyday examples of pendulums include old-fashioned clocks, a child's swing, or the sinker on a fishing line. For small displacements of less than 15 degrees, a pendulum experiences simple harmonic oscillation, meaning that its restoring force is directly proportional to its displacement. A pendulum in simple harmonic motion is called a **simple pendulum**. A pendulum has an object with a small mass, also known as the pendulum bob, which hangs from a light wire or string. The equilibrium position for a pendulum is where the angle θ is zero (that is, when the pendulum is hanging straight down). It makes sense that without any force applied, this is where the pendulum bob would rest.

The displacement of the pendulum bob is the arc length s. The weight $m\mathbf{g}$ has components $m\mathbf{g}\cos\theta$ along the string and $m\mathbf{g}\sin\theta$ tangent to the arc. Tension in the string exactly cancels the component $m\mathbf{g}\cos\theta$ parallel to the string. This leaves a *net* restoring force back toward the equilibrium position that runs tangent to the arc and equals $-m\mathbf{g}\sin\theta$.

For a simple pendulum, The period is $T = 2\pi\sqrt{\frac{L}{\mathbf{g}}}$.

The only things that affect the period of a simple pendulum are its length and the acceleration due to gravity. The period is completely independent of other factors, such as mass or amplitude. However, note that T does depend on **g**. This means that if we know the length of a pendulum, we can actually use it to measure gravity! This will come in useful in Figure 5.40.

TIPS FOR SUCCESS

Tension is represented by the variable **T**, and period is represented by the variable T. It is important not to confuse the two, since tension is a force and period is a length of time.

WORKED EXAMPLE

Measuring Acceleration due to Gravity: The Period of a Pendulum

What is the acceleration due to gravity in a region where a simple pendulum having a length 75.000 cm has a period of 1.7357 s?

Strategy

We are asked to find **g** given the period T and the length L of a pendulum. We can solve $T = 2\pi\sqrt{\frac{L}{\mathbf{g}}}$ for **g**, assuming that the angle of deflection is less than 15 degrees. Recall that when the angle of deflection is less than 15 degrees, the pendulum is considered to be in simple harmonic motion, allowing us to use this equation.

Solution

1. Square $T = 2\pi\sqrt{\frac{L}{\mathbf{g}}}$ and solve for g.

$$\mathbf{g} = 4\pi^2 \frac{L}{T^2}$$

2. Substitute known values into the new equation.

$$\mathbf{g} = 4\pi^2 \frac{0.75000 \text{ m}}{(1.7357 \text{ s})^2}$$

3. Calculate to find **g**.

$$\mathbf{g} = 9.8281 \text{ m/s}^2$$

Discussion

This method for determining **g** can be very accurate. This is why length and period are given to five digits in this example.

WORKED EXAMPLE

Hooke's Law: How Stiff Are Car Springs?

What is the force constant for the suspension system of a car, like that shown in Figure 5.41, that settles 1.20 cm when an 80.0-kg person gets in?

Figure 5.41 A car in a parking lot. (exfordy, Flickr)

Strategy

Consider the car to be in its equilibrium position $\mathbf{x} = 0$ before the person gets in. The car then settles down 1.20 cm, which means it is displaced to a position $\mathbf{x} = -1.20\times10^{-2}$ m.

At that point, the springs supply a restoring force **F** equal to the person's weight

$\mathbf{w} = m\mathbf{g} = (80.0\text{ kg})(9.80\text{ m/s}^2) = 784$ N. We take this force to be **F** in Hooke's law.

Knowing **F** and **x**, we can then solve for the force constant **k**.

Solution

Solve Hooke's law, $\mathbf{F} = -\mathbf{kx}$, for **k**.

$$\mathbf{k} = \frac{\mathbf{F}}{\mathbf{x}}$$

Substitute known values and solve for **k**.

$$\begin{aligned}\mathbf{k} &= \frac{-784\text{ N}}{-1.20\times10^{-2}\text{ m}} \\ &= 6.53\times10^{4}\text{ N/m}\end{aligned}$$

Discussion

Note that **F** and **x** have opposite signs because they are in opposite directions—the restoring force is up, and the displacement is down. Also, note that the car would oscillate up and down when the person got in, if it were not for the shock absorbers. Bouncing cars are a sure sign of bad shock absorbers.

Practice Problems

20. A force of 70 N applied to a spring causes it to be displaced by 0.3 m. What is the force constant of the spring?

a. −233 N/m
b. −21 N/m
c. 21 N/m
d. 233 N/m

21. What is the force constant for the suspension system of a car that settles 3.3 cm when a 65 kg person gets in?

a. 1.93×10^{4} N/m
b. 1.97×10^{3} N/m

c. 1.93×10^2 N/m
d. 1.97×10^1 N/m

Snap Lab

Finding Gravity Using a Simple Pendulum

Use a simple pendulum to find the acceleration due to gravity **g** in your home or classroom.

- 1 string
- 1 stopwatch
- 1 small dense object

1. Cut a piece of a string or dental floss so that it is about 1 m long.
2. Attach a small object of high density to the end of the string (for example, a metal nut or a car key).
3. Starting at an angle of less than 10 degrees, allow the pendulum to swing and measure the pendulum's period for 10 oscillations using a stopwatch.
4. Calculate **g**.

GRASP CHECK

How accurate is this measurement for *g*? How might it be improved?

a. Accuracy for value of *g* will increase with an increase in the mass of a dense object.
b. Accuracy for the value of *g* will increase with increase in the length of the pendulum.
c. The value of *g* will be more accurate if the angle of deflection is more than 15°.
d. The value of *g* will be more accurate if it maintains simple harmonic motion.

Check Your Understanding

22. What is deformation?
a. Deformation is the magnitude of the restoring force.
b. Deformation is the change in shape due to the application of force.
c. Deformation is the maximum force that can be applied on a spring.
d. Deformation is regaining the original shape upon the removal of an external force.

23. According to Hooke's law, what is deformation proportional to?
a. Force
b. Velocity
c. Displacement
d. Force constant

24. What are oscillations?
a. Motion resulting in small displacements
b. Motion which repeats itself periodically
c. Periodic, repetitive motion between two points
d. motion that is the opposite to the direction of the restoring force

25. True or False—Oscillations can occur without force.
a. True
b. False

KEY TERMS

air resistance a frictional force that slows the motion of objects as they travel through the air; when solving basic physics problems, air resistance is assumed to be zero

amplitude the maximum displacement from the equilibrium position of an object oscillating around the equilibrium position

analytical method the method of determining the magnitude and direction of a resultant vector using the Pythagorean theorem and trigonometric identities

component (of a 2-dimensional vector) a piece of a vector that points in either the vertical or the horizontal direction; every 2-d vector can be expressed as a sum of two vertical and horizontal vector components

deformation displacement from equilibrium, or change in shape due to the application of force

equilibrium position where an object would naturally rest in the absence of force

frequency number of events per unit of time

graphical method drawing vectors on a graph to add them using the head-to-tail method

head (of a vector) the end point of a vector; the location of the vector's arrow; also referred to as the tip

head-to-tail method a method of adding vectors in which the tail of each vector is placed at the head of the previous vector

Hooke's law proportional relationship between the force **F** on a material and the deformation ΔL it causes, $\mathbf{F} = \mathbf{k}\Delta L$

kinetic friction a force that opposes the motion of two systems that are in contact and moving relative to one another

maximum height (of a projectile) the highest altitude, or maximum displacement in the vertical position reached in the path of a projectile

oscillate moving back and forth regularly between two points

period time it takes to complete one oscillation

periodic motion motion that repeats itself at regular time intervals

projectile an object that travels through the air and experiences only acceleration due to gravity

projectile motion the motion of an object that is subject only to the acceleration of gravity

range the maximum horizontal distance that a projectile travels

restoring force force acting in opposition to the force caused by a deformation

resultant the sum of the a collection of vectors

resultant vector the vector sum of two or more vectors

simple harmonic motion the oscillatory motion in a system where the net force can be described by Hooke's law

simple pendulum an object with a small mass suspended from a light wire or string

static friction a force that opposes the motion of two systems that are in contact and are not moving relative to one another

tail the starting point of a vector; the point opposite to the head or tip of the arrow

trajectory the path of a projectile through the air

vector addition adding together two or more vectors

SECTION SUMMARY

5.1 Vector Addition and Subtraction: Graphical Methods

- The graphical method of adding vectors **A** and **B** involves drawing vectors on a graph and adding them by using the head-to-tail method. The resultant vector **R** is defined such that **A** + **B** = **R**. The magnitude and direction of **R** are then determined with a ruler and protractor.
- The graphical method of subtracting vectors **A** and **B** involves adding the opposite of vector **B**, which is defined as –**B**. In this case, $\mathbf{A} - \mathbf{B} = \mathbf{A} + (-\mathbf{B}) = \mathbf{R}$. Next, use the head-to-tail method as for vector addition to obtain the resultant vector **R** .
- Addition of vectors is independent of the order in which they are added; **A** + **B** = **B** + **A**.
- The head-to-tail method of adding vectors involves drawing the first vector on a graph and then placing the tail of each subsequent vector at the head of the previous vector. The resultant vector is then drawn from the tail of the first vector to the head of the final vector.
- Variables in physics problems, such as force or velocity, can be represented with vectors by making the length of the vector proportional to the magnitude of the force or velocity.
- Problems involving displacement, force, or velocity may be solved graphically by measuring the resultant vector's magnitude with a ruler and measuring the direction with a protractor.

5.2 Vector Addition and Subtraction: Analytical Methods

- The analytical method of vector addition and subtraction uses the Pythagorean theorem and trigonometric identities to determine the magnitude

and direction of a resultant vector.

- The steps to add vectors A and B using the analytical method are as follows:

1. Determine the coordinate system for the vectors. Then, determine the horizontal and vertical components of each vector using the equations

$$A_x = A\cos\theta$$
$$B_x = B\cos\theta$$

and

$$A_y = A\sin\theta$$
$$B_y = B\sin\theta.$$

2. Add the horizontal and vertical components of each vector to determine the components R_x and R_y of the resultant vector, R.

$$R_x = A_x + B_x$$

and

$$R_y = A_y + B_y.$$

3. Use the Pythagorean theorem to determine the magnitude, R, of the resultant vector R.

$$R = \sqrt{R_x^2 + R_y^2}$$

4. Use a trigonometric identity to determine the direction, θ, of R.

$$\theta = \tan^{-1}(R_y/R_x)$$

5.3 Projectile Motion

- Projectile motion is the motion of an object through the air that is subject only to the acceleration of gravity.
- Projectile motion in the horizontal and vertical directions are independent of one another.
- The maximum height of an projectile is the highest altitude, or maximum displacement in the vertical position reached in the path of a projectile.
- The range is the maximum horizontal distance traveled by a projectile.
- To solve projectile problems: choose a coordinate system; analyze the motion in the vertical and horizontal direction separately; then, recombine the horizontal and vertical components using vector addition equations.

5.4 Inclined Planes

- Friction is a contact force between systems that opposes the motion or attempted motion between them. Simple friction is proportional to the normal force **N** pushing the systems together. A normal force is always perpendicular to the contact surface between systems. Friction depends on both of the materials involved.
- μ_s is the coefficient of static friction, which depends on both of the materials.
- μ_k is the coefficient of kinetic friction, which also depends on both materials.
- When objects rest on an inclined plane that makes an angle θ with the horizontal surface, the weight of the object can be broken into components that act perpendicular ($\mathbf{w}_\perp$) and parallel ($\mathbf{w}_{||}$) to the surface of the plane.

5.5 Simple Harmonic Motion

- An oscillation is a back and forth motion of an object between two points of deformation.
- An oscillation may create a wave, which is a disturbance that propagates from where it was created.
- The simplest type of oscillations are related to systems that can be described by Hooke's law.
- Periodic motion is a repetitious oscillation.
- The time for one oscillation is the period T.
- The number of oscillations per unit time is the frequency
- A mass m suspended by a wire of length L is a simple pendulum and undergoes simple harmonic motion for amplitudes less than about 15 degrees.

KEY EQUATIONS

5.2 Vector Addition and Subtraction: Analytical Methods

resultant magnitude	$R = \sqrt{R_x^2 + R_y^2}$
resultant direction	$\theta = \tan^{-1}(R_y/R_x)$
x-component of a vector A (when an angle is given relative to the horizontal)	$A_x = A\cos\theta$
y-component of a vector A (when an angle is given relative to the horizontal)	$A_y = A\sin\theta$
addition of vectors	$\mathbf{A_x} + \mathbf{A_y} = \mathbf{A}$

5.3 Projectile Motion

angle of displacement	$\theta = \tan^{-1}(\mathbf{y}/\mathbf{x})$

velocity	$\mathbf{v} = \sqrt{\mathbf{v}_x^2 + \mathbf{v}_y^2}$
angle of velocity	$\theta_v = \tan^{-1}(\mathbf{v}_y/\mathbf{v}_x)$
maximum height	$h = \frac{\mathbf{v}_{0y}^2}{2\mathbf{g}}$
range	$R = \frac{\mathbf{v}_0^2 \sin 2\theta_0}{\mathbf{g}}$

5.4 Inclined Planes

force of static friction	$\mathbf{f}_s \leq \mu_s \mathbf{N}$
force of kinetic friction	$\mathbf{f}_k = \mu_k \mathbf{N}$
perpendicular component of weight on an inclined plane	$\mathbf{w}_\perp = \mathbf{w}\cos(\theta) = m\mathbf{g}\cos(\theta)$
parallel component of weight on an inclined plane	$\mathbf{w}_{\|\|} = \mathbf{w}\sin(\theta) = m\mathbf{g}\sin(\theta)$

5.5 Simple Harmonic Motion

Hooke's law	$\mathbf{F} = -\mathbf{kx}$
period in simple harmonic motion	$T = 2\pi\sqrt{\frac{m}{\mathbf{k}}}$
frequency in simple harmonic motion	$f = \frac{1}{2\pi}\sqrt{\frac{\mathbf{k}}{m}}$
period of a simple pendulum	$T = 2\pi\sqrt{\frac{L}{\mathbf{g}}}$

CHAPTER REVIEW

Concept Items

5.1 Vector Addition and Subtraction: Graphical Methods

1. There is a vector $\vec{A}$, with magnitude 5 units pointing towards west and vector $\vec{B}$, with magnitude 3 units, pointing towards south. Using vector addition, calculate the magnitude of the resultant vector.
   a. 4.0
   b. 5.8
   c. 6.3
   d. 8.0

2. If you draw two vectors using the head-to-tail method, how can you then draw the resultant vector?
   a. By joining the head of the first vector to the head of the last
   b. By joining the head of the first vector with the tail of the last
   c. By joining the tail of the first vector to the head of the last
   d. By joining the tail of the first vector with the tail of the last

3. What is the global angle of 20° south of west?
   a. 110°
   b. 160°
   c. 200°
   d. 290°

5.2 Vector Addition and Subtraction: Analytical Methods

4. What is the angle between the x and y components of a vector?
   a. 0°
   b. 45°
   c. 90°
   d. 180°

5. Two vectors are equal in magnitude and opposite in direction. What is the magnitude of their resultant vector?
   a. The magnitude of the resultant vector will be zero.
   b. The magnitude of resultant vector will be twice the magnitude of the original vector.
   c. The magnitude of resultant vector will be same as magnitude of the original vector.
   d. The magnitude of resultant vector will be half the magnitude of the original vector.

6. How can we express the x and y-components of a vector in terms of its magnitude, A, and direction, global angle θ?
   a. $A_x = A\cos\theta\, A_y = A\sin\theta$
   b. $A_x = A\cos\theta\, A_y = A\cos\theta$
   c. $A_x = A\sin\theta\, A_y = A\cos\theta$

d. $A_x = A \sin\theta A_y = A \sin\theta$

7. True or False—Every 2-D vector can be expressed as the product of its x and y-components.
   a. True
   b. False

5.3 Projectile Motion

8. Horizontal and vertical motions of a projectile are independent of each other. What is meant by this?
   a. Any object in projectile motion falls at the same rate as an object in freefall, regardless of its horizontal velocity.
   b. All objects in projectile motion fall at different rates, regardless of their initial horizontal velocities.
   c. Any object in projectile motion falls at the same rate as its initial vertical velocity, regardless of its initial horizontal velocity.
   d. All objects in projectile motion fall at different rates and the rate of fall of the object is independent of the initial velocity.

9. Using the conventional choice for positive and negative axes described in the text, what is the y-component of the acceleration of an object experiencing projectile motion?
   a. $-9.8\ \text{m/s}$
   b. $-9.8\ \text{m/s}^2$
   c. $9.8\ \text{m/s}$
   d. $9.8\ \text{m/s}^2$

5.4 Inclined Planes

10. True or False—Kinetic friction is less than the limiting static friction because once an object is moving, there are fewer points of contact, and the friction is reduced. For this reason, more force is needed to start moving an object than to keep it in motion.
   a. True
   b. False

11. When there is no motion between objects, what is the relationship between the magnitude of the static friction f_s and the normal force N?
   a. $f_s \leq N$
   b. $f_s \leq \mu_s N$
   c. $f_s \geq N$
   d. $f_s \geq \mu_s N$

12. What equation gives the magnitude of kinetic friction?
   a. $f_k = \mu_s N$
   b. $f_k = \mu_k N$
   c. $f_k \leq \mu_s N$
   d. $f_k \leq \mu_k N$

5.5 Simple Harmonic Motion

13. Why is there a negative sign in the equation for Hooke's law?
   a. The negative sign indicates that displacement decreases with increasing force.
   b. The negative sign indicates that the direction of the applied force is opposite to that of displacement.
   c. The negative sign indicates that the direction of the restoring force is opposite to that of displacement.
   d. The negative sign indicates that the force constant must be negative.

14. With reference to simple harmonic motion, what is the equilibrium position?
   a. The position where velocity is the minimum
   b. The position where the displacement is maximum
   c. The position where the restoring force is the maximum
   d. The position where the object rests in the absence of force

15. What is Hooke's law?
   a. Restoring force is directly proportional to the displacement from the mean position and acts in the the opposite direction of the displacement.
   b. Restoring force is directly proportional to the displacement from the mean position and acts in the same direction as the displacement.
   c. Restoring force is directly proportional to the square of the displacement from the mean position and acts in the opposite direction of the displacement.
   d. Restoring force is directly proportional to the square of the displacement from the mean position and acts in the same direction as the displacement.

Critical Thinking Items

5.1 Vector Addition and Subtraction: Graphical Methods

16. True or False—A person is following a set of directions. He has to walk 2 km east and then 1 km north. He takes a wrong turn and walks in the opposite direction for the second leg of the trip. The magnitude of his total displacement will be the same as it would have been had he followed directions correctly.
   a. True
   b. False

5.2 Vector Addition and Subtraction: Analytical Methods

17. What is the magnitude of a vector whose x-component is 2 units and whose angle is 60°?
a. 1.0 units
b. 2.0 units
c. 2.3 units
d. 4.0 units

18. Vectors $\vec{A}$ and $\vec{B}$ are equal in magnitude and opposite in direction. Does $\vec{A} - \vec{B}$ have the same direction as vector $\vec{A}$ or $\vec{B}$?
a. $\vec{A}$
b. $\vec{B}$

5.3 Projectile Motion

19. Two identical items, object 1 and object 2, are dropped from the top of a 50.0 m building. Object 1 is dropped with an initial velocity of 0 m/s, while object 2 is thrown straight downward with an initial velocity of 13.0 m/s. What is the difference in time, in seconds rounded to the nearest tenth, between when the two objects hit the ground?
a. Object 1 will hit the ground 3.2 s after object 2.
b. Object 1 will hit the ground 2.1 s after object 2.
c. Object 1 will hit the ground at the same time as object 2.
d. Object 1 will hit the ground 1.1 s after object 2.

20. An object is launched into the air. If the y-component of its acceleration is 9.8 m/s^2, which direction is defined as positive?
a. Vertically upward in the coordinate system
b. Vertically downward in the coordinate system
c. Horizontally to the right side of the coordinate system
d. Horizontally to the left side of the coordinate system

5.4 Inclined Planes

21. A box weighing 500 N is at rest on the floor. A person pushes against it and it starts moving when 100 N force is applied to it. What can be said about the coefficient of kinetic friction between the box and the floor?
a. $\mu_k = 0$
b. $\mu_k = 0.2$
c. $\mu_k < 0.2$
d. $\mu_k > 0.2$

22. The component of the weight parallel to an inclined plane of an object resting on an incline that makes an angle of 70.0° with the horizontal is 100.0 N. What is the object's mass?
a. 10.9 kg
b. 29.8 kg
c. 106 kg
d. 292 kg

5.5 Simple Harmonic Motion

23. Two springs are attached to two hooks. Spring A has a greater force constant than spring B. Equal weights are suspended from both. Which of the following statements is true?
a. Spring A will have more extension than spring B.
b. Spring B will have more extension than spring A.
c. Both springs will have equal extension.
d. Both springs are equally stiff.

24. Two simple harmonic oscillators are constructed by attaching similar objects to two different springs. The force constant of the spring on the left is 5 N/m and that of the spring on the right is 4 N/m. If the same force is applied to both, which of the following statements is true?
a. The spring on the left will oscillate faster than spring on the right.
b. The spring on the right will oscillate faster than the spring on the left.
c. Both the springs will oscillate at the same rate.
d. The rate of oscillation is independent of the force constant.

Problems

5.1 Vector Addition and Subtraction: Graphical Methods

25. A person attempts to cross a river in a straight line by navigating a boat at 15 m/s. If the river flows at 5.0 m/s from his left to right, what would be the magnitude of the boat's resultant velocity? In what direction would the boat go, relative to the straight line across it?
a. The resultant velocity of the boat will be 10.0 m/s. The boat will go toward his right at an angle of 26.6° to a line drawn across the river.
b. The resultant velocity of the boat will be 10.0 m/s. The boat will go toward his left at an angle of 26.6° to a line drawn across the river.
c. The resultant velocity of the boat will be 15.8 m/s. The boat will go toward his right at an angle of

18.4° to a line drawn across the river.

d. The resultant velocity of the boat will be $15.8\ \text{m/s}$. The boat will go toward his left at an angle of 18.4° to a line drawn across the river.

26. A river flows in a direction from south west to north east at a velocity of $7.1\ \text{m/s}$. A boat captain wants to cross this river to reach a point on the opposite shore due east of the boat's current position. The boat moves at $13\ \text{m/s}$. Which direction should it head towards if the resultant velocity is $19.74\ \text{m/s}$?

a. It should head in a direction 22.6° east of south.
b. It should head in a direction 22.6° south of east.
c. It should head in a direction 45.0° east of south.
d. It should head in a direction 45.0° south of east.

5.2 Vector Addition and Subtraction: Analytical Methods

27. A person walks 10.0 m north and then 2.00 m east. Solving analytically, what is the resultant displacement of the person?

a. $\vec{R} = 10.2\ \text{m}, \theta = 78.7^\circ$ east of north
b. $\vec{R} = 10.2\ \text{m}, \theta = 78.7^\circ$ north of east
c. $\vec{R} = 12.0\ \text{m}, \theta = 78.7^\circ$ east of north
d. $\vec{R} = 12.0\ \text{m}, \theta = 78.7^\circ$ north of east

28. A person walks 12.0° north of west for $55.0\ \text{m}$ and 63.0° south of west for $25.0\ \text{m}$. What is the magnitude of his displacement? Solve analytically.

a. $10.84\ \text{m}$
b. $65.1\ \text{m}$
c. $66.04\ \text{m}$
d. $80.00\ \text{m}$

5.3 Projectile Motion

29. A water balloon cannon is fired at $30\ \text{m/s}$ at an angle of 50° above the horizontal. How far away will it fall?

a. $2.35\ \text{m}$
b. $3.01\ \text{m}$
c. $70.35\ \text{m}$
d. $90.44\ \text{m}$

30. A person wants to fire a water balloon cannon such that it hits a target 100 m away. If the cannon can only be launched at 45° above the horizontal, what should be the initial speed at which it is launched?

a. 31.3 m/s
b. 37.2 m/s
c. 980.0 m/s
d. 1,385.9 m/s

5.4 Inclined Planes

31. A coin is sliding down an inclined plane at constant velocity. If the angle of the plane is 10° to the horizontal, what is the coefficient of kinetic friction?

a. $\mu_k = 0$
b. $\mu_k = 0.18$
c. $\mu_k = 5.88$
d. $\mu_k = \infty$

32. A skier with a mass of 55 kg is skiing down a snowy slope that has an incline of 30°. Find the coefficient of kinetic friction for the skier if friction is known to be 25 N .

a. $\mu k = 0$
b. $\mu k = 0.05$
c. $\mu k = 0.09$
d. $\mu k = \infty$

5.5 Simple Harmonic Motion

33. What is the time period of a $6\ \text{cm}$ long pendulum on earth?

a. $0.08\ \text{s}$
b. $0.49\ \text{s}$
c. $4.9\ \text{s}$
d. $80\ \text{s}$

34. A simple harmonic oscillator has time period $4\ \text{s}$. If the mass of the system is $2\ \text{kg}$, what is the force constant of the spring used?

a. $0.125\ \text{N/m}$
b. $0.202\ \text{N/m}$
c. $0.81\ \text{N/m}$
d. $4.93\ \text{N/m}$

Performance Task

5.5 Simple Harmonic Motion

35. Construct a seconds pendulum (pendulum with time period 2 seconds).

TEST PREP

Multiple Choice

5.1 Vector Addition and Subtraction: Graphical Methods

36. True or False—We can use Pythagorean theorem to calculate the length of the resultant vector obtained from the addition of two vectors which are at right angles to each other.
a. True
b. False

37. True or False—The direction of the resultant vector depends on both the magnitude and direction of added vectors.
a. True
b. False

38. A plane flies north at 200 m/s with a headwind blowing from the north at 70 m/s. What is the resultant velocity of the plane?
a. 130 m/s north
b. 130 m/s south
c. 270 m/s north
d. 270 m/s south

39. Two hikers take different routes to reach the same spot. The first one goes 255 m southeast, then turns and goes 82 m at 14° south of east. The second hiker goes 200 m south. How far and in which direction must the second hiker travel now, in order to reach the first hiker's location destination?
a. 200 m east
b. 200 m south
c. 260 m east
d. 260 m south

5.2 Vector Addition and Subtraction: Analytical Methods

40. When will the x-component of a vector with angle θ be greater than its y-component?
a. $0^\circ < \theta < 45^\circ$
b. $\theta = 45^\circ$
c. $45^\circ < \theta < 60^\circ$
d. $60^\circ < \theta < 90^\circ$

41. The resultant vector of the addition of vectors $\vec{a}$ and $\vec{b}$ is $\vec{r}$. The magnitudes of $\vec{a}$, $\vec{b}$, and $\vec{r}$ are A, B, and R, respectively. Which of the following is true?
a. $R_x + R_y = 0$
b. $A_x + A_y = \vec{A}$
c. $A_x + B_y = B_x + A_y$
d. $A_x + B_x = R_x$

42. What is the dimensionality of vectors used in the study of atmospheric sciences?
a. One-dimensional
b. Two-dimensional
c. Three-dimensional

5.3 Projectile Motion

43. After a projectile is launched in the air, in which direction does it experience constant, non-zero acceleration, ignoring air resistance?
a. The x direction
b. The y direction
c. Both the x and y directions
d. Neither direction

44. Which is true when the height of a projectile is at its maximum?
a. $v_y = 0$
b. $v_y = \text{maximum}$
c. $v_x = \text{maximum}$

45. A ball is thrown in the air at an angle of 40°. If the maximum height it reaches is 10 m, what must be its initial speed?
a. 17.46 m/s
b. 21.78 m/s
c. 304.92 m/s
d. 474.37 m/s

46. A large rock is ejected from a volcano with a speed of 30 m/s and at an angle 60° above the horizontal. The rock strikes the side of the volcano at an altitude of 10.0 m lower than its starting point. Calculate the horizontal displacement of the rock.
a. 84.90 m
b. 96.59 m
c. 169.80 m
d. 193.20 m

5.4 Inclined Planes

47. For objects of identical masses but made of different materials, which of the following experiences the most static friction?
a. Shoes on ice
b. Metal on wood
c. Teflon on steel

48. If an object sits on an inclined plane and no other object makes contact with the object, what is typically equal in magnitude to the component of the weight perpendicular to the plane?

a. The normal force
b. The total weight
c. The parallel force of weight

49. A 5 kg box is at rest on the floor. The coefficient of static friction between the box and the floor is 0.4. A horizontal force of 50 N is applied to the box. Will it move?
a. No, because the applied force is less than the maximum limiting static friction.
b. No, because the applied force is more than the maximum limiting static friction.
c. Yes, because the applied force is less than the maximum limiting static friction.
d. Yes, because the applied force is more than the maximum limiting static friction.

50. A skier with a mass of 67 kg is skiing down a snowy slope with an incline of 37°. Find the friction if the coefficient of kinetic friction is 0.07.
a. 27.66 N
b. 34.70 N
c. 36.71 N
d. 45.96 N

5.5 Simple Harmonic Motion

51. A change in which of the following is an example of deformation?
a. Velocity
b. Length
c. Mass
d. Weight

52. The units of amplitude are the same as those for which of the following measurements?
a. Speed
b. Displacement
c. Acceleration
d. Force

53. Up to approximately what angle is simple harmonic motion a good model for a pendulum?
a. 15°
b. 45°
c. 75°
d. 90°

54. How would simple harmonic motion be different in the absence of friction?
a. Oscillation will not happen in the absence of friction.
b. Oscillation will continue forever in the absence of friction.
c. Oscillation will have changing amplitude in the absence of friction.
d. Oscillation will cease after a certain amount of time in the absence of friction.

55. What mass needs to be attached to a spring with a force constant of $7\ \mathrm{N/m}$ in order to make a simple harmonic oscillator oscillate with a time period of 3 s?
a. $0.03\ \mathrm{kg}$
b. $1.60\ \mathrm{kg}$
c. $30.7\ \mathrm{kg}$
d. $63.0\ \mathrm{kg}$

Short Answer

5.1 Vector Addition and Subtraction: Graphical Methods

56. Find $\vec{A} - \vec{B}$ for the following vectors:
$\vec{A} = (122\ \mathrm{cm}, \angle 145^\circ)\ \vec{B} = (110\ \mathrm{cm}, \angle 270^\circ)$
a. 108 cm, $\theta = 119.0^\circ$
b. 108 cm, $\theta = 125.0^\circ$
c. 206 cm, $\theta = 119.0^\circ$
d. 206 cm, $\theta = 125.0^\circ$

57. Find $\vec{A} + \vec{B}$ for the following vectors:
$\vec{A} = (122\ \mathrm{cm}, \angle 145^\circ)\ \vec{B} = (110\ \mathrm{cm}, \angle 270^\circ)$
a. 108 cm, $\theta = 119.1^\circ$
b. 108 cm, $\theta = 201.8^\circ$
c. 232 cm, $\theta = 119.1^\circ$
d. 232 cm, $\theta = 201.8^\circ$

58. Consider six vectors of 2 cm each, joined from head to tail making a hexagon. What would be the magnitude of the addition of these vectors?
a. Zero
b. Six
c. Eight
d. Twelve

59. Two people pull on ropes tied to a trolley, each applying 44 N of force. The angle the ropes form with each other is 39.5°. What is the magnitude of the net force exerted on the trolley?
a. 0.0 N
b. 79.6 N
c. 82.8 N
d. 88.0 N

5.2 Vector Addition and Subtraction: Analytical Methods

60. True or False—A vector can form the shape of a right angle triangle with its x and y components.
a. True

b. False

61. True or False—All vectors have positive x and y components.
   a. True
   b. False

62. Consider $\vec{A} - \vec{B} = \vec{R}$. What is R_x in terms of A_x and B_x?
   a. $R_x = \frac{A_x}{B_x}$
   b. $R_x = \frac{B_x}{A_x}$
   c. $R_x = A_x + B_x$
   d. $R_x = A_x - B_x$

63. Consider $\vec{A} - \vec{B} = \vec{R}$. What is R_y in terms of A_y and B_y?
   a. $R_y = \frac{A_y}{B_y}$
   b. $R_y = \frac{B_y}{A_y}$
   c. $R_y = A_y + B_y$
   d. $R_y = A_y - B_y$

64. When a three dimensional vector is used in the study of atmospheric sciences, what is z?
   a. Altitude
   b. Heat
   c. Temperature
   d. Wind speed

65. Which method is not an application of vector calculus?
   a. To find the rate of change in atmospheric temperature
   b. To study changes in wind speed and direction
   c. To predict changes in atmospheric pressure
   d. To measure changes in average rainfall

5.3 Projectile Motion

66. How can you express the velocity, $\vec{v}$, of a projectile in terms of its initial velocity, $\vec{v_0}$, acceleration, $\vec{a}$, and time, t?
   a. $\vec{v} = \vec{a}t$
   b. $\vec{v} = \vec{v_0} + \vec{a}t$
   c. $\vec{v} + \vec{v_0} = \vec{a}t$
   d. $\vec{v_0} + \vec{v} + \vec{a}t$

67. In the equation for the maximum height of a projectile, what does v_{0y} stand for? $h = \frac{{v_{0y}}^2}{2g}$
   a. Initial velocity in the x direction
   b. Initial velocity in the y direction
   c. Final velocity in the x direction
   d. Final velocity in the y direction

68. True or False—Range is defined as the maximum vertical distance travelled by a projectile.
   a. True
   b. False

69. For what angle of a projectile is its range equal to zero?
   a. 0° or 30°
   b. 0° or 45°
   c. 90° or 0°
   d. 90° or 45°

5.4 Inclined Planes

70. What are the units of the coefficient of friction?
   a. N
   b. m/s
   c. m/s^2
   d. unitless

71. Two surfaces in contact are moving slowly past each other. As the relative speed between the two surfaces in contact increases, what happens to the magnitude of their coefficient of kinetic friction?
   a. It increases with the increase in the relative motion.
   b. It decreases with the increase in the relative motion.
   c. It remains constant and is independent of the relative motion.

72. When will an object slide down an inclined plane at constant velocity?
   a. When the magnitude of the component of the weight along the slope is equal to the magnitude of the frictional force.
   b. When the magnitude of the component of the weight along the slope is greater than the magnitude of the frictional force.
   c. When the magnitude of the component of the weight perpendicular to the slope is less than the magnitude of the frictional force.
   d. When the magnitude of the component of the weight perpendicular to the slope is equal to the magnitude of the frictional force.

73. A box is sitting on an inclined plane. At what angle of incline is the perpendicular component of the box's weight at its maximum?
   a. 0°
   b. 30°
   c. 60°
   d. 90°

5.5 Simple Harmonic Motion

74. What is the term used for changes in shape due to the application of force?
   a. Amplitude

b. Deformation
c. Displacement
d. Restoring force

75. What is the restoring force?
a. The normal force on the surface of an object
b. The weight of a mass attached to an object
c. Force which is applied to deform an object from its original shape
d. Force which brings an object back to its equilibrium position

76. For a given oscillator, what are the factors that affect its period and frequency?
a. Mass only
b. Force constant only
c. Applied force and mass
d. Mass and force constant

77. For an object in simple harmonic motion, when does the maximum speed occur?
a. At the extreme positions
b. At the equilibrium position
c. At the moment when the applied force is removed
d. Midway between the extreme and equilibrium positions

78. What is the equilibrium position of a pendulum?
a. When the tension in the string is zero
b. When the pendulum is hanging straight down
c. When the tension in the string is maximum
d. When the weight of the mass attached is minimum

79. If a pendulum is displaced by an angle θ, what is the net restoring force it experiences?
a. $mg\sin\theta$
b. $mg\cos\theta$
c. $-mg\sin\theta$
d. $-mg\cos\theta$

Extended Response

5.1 Vector Addition and Subtraction: Graphical Methods

80. True or False—For vectors the order of addition is important.
a. True
b. False

81. Consider five vectors *a*, *b*, *c*, *d*, and *e*. Is it true or false that their addition always results in a vector with a greater magnitude than if only two of the vectors were added?
a. True
b. False

5.2 Vector Addition and Subtraction: Analytical Methods

82. For what angle of a vector is it possible that its magnitude will be equal to its y-component?
a. $\theta = 0^\circ$
b. $\theta = 45^\circ$
c. $\theta = 60^\circ$
d. $\theta = 90^\circ$

83. True or False—If only the angles of two vectors are known, we can find the angle of their resultant addition vector.
a. True
b. False

84. True or false—We can find the magnitude and direction of the resultant vector if we know the angles of two vectors and the magnitude of one.
a. True
b. False

5.3 Projectile Motion

85. Ignoring drag, what is the x-component of the acceleration of a projectile? Why?
a. The x-component of the acceleration of a projectile is 0 because acceleration of a projectile is due to gravity, which acts in the y direction.
b. The x component of the acceleration of a projectile is g because acceleration of a projectile is due to gravity, which acts in the x direction.
c. The x-component of the acceleration of a projectile is 0 because acceleration of a projectile is due to gravity, which acts in the x direction.
d. The x-component of the acceleration of a projectile is g because acceleration of a projectile is due to gravity, which acts in the y direction.

86. What is the optimum angle at which a projectile should be launched in order to cover the maximum distance?
a. 0°
b. 45°
c. 60°
d. 90°

5.4 Inclined Planes

87. True or False—Friction varies from surface to surface because different substances have different degrees of roughness or smoothness.
a. True
b. False

88. As the angle of the incline gets larger, what happens to

the magnitudes of the perpendicular and parallel components of gravitational force?

a. Both the perpendicular and the parallel component will decrease.
b. The perpendicular component will decrease and the parallel component will increase.
c. The perpendicular component will increase and the parallel component will decrease.
d. Both the perpendicular and the parallel component will increase.

5.5 Simple Harmonic Motion

89. What physical characteristic of a system is its force constant related to?

a. The force constant *k* is related to the stiffness of a system: The larger the force constant, the stiffer the system.
b. The force constant *k* is related to the stiffness of a system: The larger the force constant, the looser the system.
c. The force constant *k* is related to the friction in the system: The larger the force constant, the greater the friction in the system.
d. The force constant *k* is related to the friction in the system: The larger the force constant, the lower the friction in the system.

90. How or why does a pendulum oscillate?

a. A pendulum oscillates due to applied force.
b. A pendulum oscillates due to the elastic nature of the string.
c. A pendulum oscillates due to restoring force arising from gravity.
d. A pendulum oscillates due to restoring force arising from tension in the string.

91. If a pendulum from earth is taken to the moon, will its frequency increase or decrease? Why?

a. It will increase because *g* on the Moon is less than *g* on Earth.
b. It will decrease because *g* on the Moon is less than *g* on Earth.
c. It will increase because *g* on the Moon is greater than *g* on Earth.
d. It will decrease because *g* on the Moon is greater than *g* on Earth.

CHAPTER 6
Circular and Rotational Motion

Figure 6.1 This Australian Grand Prix Formula 1 race car moves in a circular path as it makes the turn. Its wheels also spin rapidly. The same physical principles are involved in both of these motions. (Richard Munckton).

Chapter Outline

6.1 Angle of Rotation and Angular Velocity

6.2 Uniform Circular Motion

6.3 Rotational Motion

INTRODUCTION You may recall learning about various aspects of motion along a straight line: kinematics (where we learned about displacement, velocity, and acceleration), projectile motion (a special case of two-dimensional kinematics), force, and Newton's laws of motion. In some ways, this chapter is a continuation of Newton's laws of motion. Recall that Newton's first law tells us that objects move along a straight line at constant speed unless a net external force acts on them. Therefore, if an object moves along a circular path, such as the car in the photo, it must be experiencing an external force. In this chapter, we explore both circular motion and rotational motion.

6.1 Angle of Rotation and Angular Velocity

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Describe the angle of rotation and relate it to its linear counterpart
- Describe angular velocity and relate it to its linear counterpart
- Solve problems involving angle of rotation and angular velocity

Section Key Terms

angle of rotation	angular velocity	arc length	circular motion
radius of curvature	rotational motion	spin	tangential velocity

Angle of Rotation

What exactly do we mean by *circular motion* or *rotation*? **Rotational motion** is the circular motion of an object about an axis of rotation. We will discuss specifically circular motion and spin. **Circular motion** is when an object moves in a circular path. Examples of circular motion include a race car speeding around a circular curve, a toy attached to a string swinging in a circle around your head, or the circular *loop-the-loop* on a roller coaster. **Spin** is rotation about an axis that goes through the center of mass of the object, such as Earth rotating on its axis, a wheel turning on its axle, the spin of a tornado on its path of destruction, or a figure skater spinning during a performance at the Olympics. Sometimes, objects will be spinning while in circular motion, like the Earth spinning on its axis while revolving around the Sun, but we will focus on these two motions separately.

When solving problems involving rotational motion, we use variables that are similar to linear variables (distance, velocity, acceleration, and force) but take into account the curvature or rotation of the motion. Here, we define the **angle of rotation**, which is the angular equivalence of distance; and **angular velocity**, which is the angular equivalence of linear velocity.

When objects rotate about some axis—for example, when the CD in Figure 6.2 rotates about its center—each point in the object follows a circular path.

Figure 6.2 All points on a CD travel in circular paths. The pits (dots) along a line from the center to the edge all move through the same angle $\Delta\theta$ in time Δt .

The **arc length,** , is the distance traveled along a circular path. The **radius of curvature**, **r**, is the radius of the circular path. Both are shown in Figure 6.3.

Figure 6.3 The radius (r) of a circle is rotated through an angle $\Delta\theta$. The arc length, Δs, is the distance covered along the circumference.

Consider a line from the center of the CD to its edge. In a given time, each *pit* (used to record information) on this line moves through the same angle. The angle of rotation is the amount of rotation and is the angular analog of distance. The angle of rotation $\Delta\theta$ is the arc length divided by the radius of curvature.

$$\Delta\theta = \frac{\Delta s}{r}$$

The angle of rotation is often measured by using a unit called the radian. (Radians are actually dimensionless, because a radian is defined as the ratio of two distances, radius and arc length.) A revolution is one complete rotation, where every point on the circle returns to its original position. One revolution covers 2π radians (or 360 degrees), and therefore has an angle of rotation of 2π radians, and an arc length that is the same as the circumference of the circle. We can convert between radians, revolutions, and degrees using the relationship

1 revolution = 2π rad = 360°. See Table 6.1 for the conversion of degrees to radians for some common angles.

$$\begin{aligned} 2\pi \text{ rad} &= 360^\circ \\ 1 \text{ rad} &= \frac{360^\circ}{2\pi} \approx 57.3^\circ \end{aligned} \tag{6.1}$$

Degree Measures	Radian Measures
30°	$\frac{\pi}{6}$
60°	$\frac{\pi}{3}$
90°	$\frac{\pi}{2}$
120°	$\frac{2\pi}{3}$
135°	$\frac{3\pi}{4}$
180°	π

Table 6.1 Commonly Used Angles in Terms of Degrees and Radians

Angular Velocity

How fast is an object rotating? We can answer this question by using the concept of angular velocity. Consider first the angular speed (ω) is the rate at which the angle of rotation changes. In equation form, the angular speed is

$$\omega = \frac{\Delta\theta}{\Delta t}, \tag{6.2}$$

which means that an angular rotation ($\Delta\theta$) occurs in a time, Δt. If an object rotates through a greater angle of rotation in a given time, it has a greater angular speed. The units for angular speed are radians per second (rad/s).

Now let's consider the direction of the angular speed, which means we now must call it the angular velocity. The direction of the

angular velocity is along the axis of rotation. For an object rotating clockwise, the angular velocity points away from you along the axis of rotation. For an object rotating counterclockwise, the angular velocity points toward you along the axis of rotation.

Angular velocity (ω) is the angular version of linear velocity **v**. **Tangential velocity** is the instantaneous linear velocity of an object in rotational motion**.** To get the precise relationship between angular velocity and tangential velocity, consider again a pit on the rotating CD. This pit moves through an arc length (Δs) in a *short* time (Δt) so its tangential *speed* is

$$v = \frac{\Delta s}{\Delta t}. \tag{6.3}$$

From the definition of the angle of rotation, $\Delta\theta = \frac{\Delta s}{r}$, we see that $\Delta s = r\Delta\theta$. Substituting this into the expression for v gives

$$v = \frac{r\Delta\theta}{\Delta t} = r\omega.$$

The equation $v = r\omega$ says that the tangential speed v is proportional to the distance r from the center of rotation. Consequently, tangential speed is greater for a point on the outer edge of the CD (with larger r) than for a point closer to the center of the CD (with smaller r). This makes sense because a point farther out from the center has to cover a longer arc length in the same amount of time as a point closer to the center. Note that both points will still have the same angular speed, regardless of their distance from the center of rotation. See Figure 6.4.

Figure 6.4 Points 1 and 2 rotate through the same angle ($\Delta\theta$), but point 2 moves through a greater arc length (Δs_2) because it is farther from the center of rotation.

Now, consider another example: the tire of a moving car (see Figure 6.5). The faster the tire spins, the faster the car moves—large ω means large v because $v = r\omega$. Similarly, a larger-radius tire rotating at the same angular velocity, **ω**, will produce a greater linear (tangential) velocity, **v,** for the car. This is because a larger radius means a longer arc length must contact the road, so the car must move farther in the same amount of time.

Figure 6.5 A car moving at a velocity, **v,** to the right has a tire rotating with angular velocity **ω** . The speed of the tread of the tire relative to the axle is *v*, the same as if the car were jacked up and the wheels spinning without touching the road. Directly below the axle, where the tire touches the road, the tire tread moves backward with respect to the axle with tangential velocity $v = r\omega$, where *r* is the tire radius. Because the road is stationary with respect to this point of the tire, the car must move forward at the linear velocity **v**. A larger angular velocity for the tire means a greater linear velocity for the car.

However, there are cases where linear velocity and tangential velocity are not equivalent, such as a car spinning its tires on ice. In this case, the linear velocity will be less than the tangential velocity. Due to the lack of friction under the tires of a car on ice, the arc length through which the tire treads move is greater than the linear distance through which the car moves. It's similar to running on a treadmill or pedaling a stationary bike; you are literally going nowhere fast.

TIPS FOR SUCCESS

Angular velocity **ω** and tangential velocity **v** are vectors, so we must include magnitude and direction. The direction of the angular velocity is along the axis of rotation, and points away from you for an object rotating clockwise, and toward you for an object rotating counterclockwise. In mathematics this is described by the right-hand rule. Tangential velocity is usually described as up, down, left, right, north, south, east, or west, as shown in Figure 6.6.

Figure 6.6 As the fly on the edge of an old-fashioned vinyl record moves in a circle, its instantaneous velocity is always at a tangent to the circle. The direction of the angular velocity is into the page this case.

WATCH PHYSICS

Relationship between Angular Velocity and Speed

This video reviews the definition and units of angular velocity and relates it to linear speed. It also shows how to convert between revolutions and radians.

Click to view content (https://www.youtube.com/embed/zAx61CO5mDw)

GRASP CHECK

For an object traveling in a circular path at a constant angular speed, would the linear speed of the object change if the radius of the path increases?

a. Yes, because tangential speed is independent of the radius.
b. Yes, because tangential speed depends on the radius.
c. No, because tangential speed is independent of the radius.
d. No, because tangential speed depends on the radius.

Solving Problems Involving Angle of Rotation and Angular Velocity

Snap Lab

Measuring Angular Speed

In this activity, you will create and measure uniform circular motion and then contrast it with circular motions with different radii.

- One string (1 m long)
- One object (two-hole rubber stopper) to tie to the end
- One timer

Procedure

1. Tie an object to the end of a string.
2. Swing the object around in a horizontal circle above your head (swing from your wrist). It is important that the circle be horizontal!
3. Maintain the object at uniform speed as it swings.
4. Measure the angular speed of the object in this manner. Measure the time it takes in seconds for the object to travel 10 revolutions. Divide that time by 10 to get the angular speed in revolutions per second, which you can convert to radians per second.
5. What is the approximate linear speed of the object?
6. Move your hand up the string so that the length of the string is 90 cm. Repeat steps 2–5.
7. Move your hand up the string so that its length is 80 cm. Repeat steps 2–5.
8. Move your hand up the string so that its length is 70 cm. Repeat steps 2–5.
9. Move your hand up the string so that its length is 60 cm. Repeat steps 2–5
10. Move your hand up the string so that its length is 50 cm. Repeat steps 2–5
11. Make graphs of angular speed vs. radius (i.e. string length) and linear speed vs. radius. Describe what each graph looks like.

GRASP CHECK

If you swing an object slowly, it may rotate at less than one revolution per second. What would be the revolutions per second for an object that makes one revolution in five seconds? What would be its angular speed in radians per second?

a. The object would spin at $\frac{1}{5}$ rev/s. The angular speed of the object would be $\frac{2\pi}{5}$ rad/s.
b. The object would spin at $\frac{1}{5}$ rev/s. The angular speed of the object would be $\frac{\pi}{5}$ rad/s.

c. The object would spin at 5 rev/s. The angular speed of the object would be 10π rad/s.
d. The object would spin at 5 rev/s. The angular speed of the object would be 5π rad/s.

Now that we have an understanding of the concepts of angle of rotation and angular velocity, we'll apply them to the real-world situations of a clock tower and a spinning tire.

WORKED EXAMPLE

Angle of rotation at a Clock Tower

The clock on a clock tower has a radius of 1.0 m. (a) What angle of rotation does the hour hand of the clock travel through when it moves from 12 p.m. to 3 p.m.? (b) What's the arc length along the outermost edge of the clock between the hour hand at these two times?

Strategy

We can figure out the angle of rotation by multiplying a full revolution (2π radians) by the fraction of the 12 hours covered by the hour hand in going from 12 to 3. Once we have the angle of rotation, we can solve for the arc length by rearranging the equation $\Delta\theta = \frac{\Delta s}{r}$ since the radius is given.

Solution to (a)

In going from 12 to 3, the hour hand covers 1/4 of the 12 hours needed to make a complete revolution. Therefore, the angle between the hour hand at 12 and at 3 is $\frac{1}{4} \times 2\ \pi\text{rad} = \frac{\pi}{2}$ (i.e., 90 degrees).

Solution to (b)

Rearranging the equation

$$\Delta\theta = \frac{\Delta s}{r}, \tag{6.4}$$

we get

$$\Delta s = r\Delta\theta. \tag{6.5}$$

Inserting the known values gives an arc length of

$$\begin{aligned} \Delta s &= (1.0\text{ m})\left(\frac{\pi}{2}\text{rad}\right) \\ &= 1.6\text{ m} \end{aligned} \tag{6.6}$$

Discussion

We were able to drop the radians from the final solution to part (b) because radians are actually dimensionless. This is because the radian is defined as the ratio of two distances (radius and arc length). Thus, the formula gives an answer in units of meters, as expected for an arc length.

WORKED EXAMPLE

How Fast Does a Car Tire Spin?

Calculate the angular speed of a 0.300 m radius car tire when the car travels at 15.0 m/s (about 54 km/h). See Figure 6.5.

Strategy

In this case, the speed of the tire tread with respect to the tire axle is the same as the speed of the car with respect to the road, so we have $v = 15.0$ m/s. The radius of the tire is $r = 0.300$ m. Since we know v and r, we can rearrange the equation $v = r\omega$, to get $\omega = \frac{v}{r}$ and find the angular speed.

Solution

To find the angular speed, we use the relationship: $\omega = \frac{v}{r}$.

Inserting the known quantities gives

$$\begin{aligned}\omega &= \frac{15.0\ \text{m/s}}{0.300\ \text{m}} \\ &= 50.0\ \text{rad/s}.\end{aligned} \tag{6.7}$$

Discussion

When we cancel units in the above calculation, we get 50.0/s (i.e., 50.0 per second, which is usually written as 50.0 s^{-1}). But the angular speed must have units of rad/s. Because radians are dimensionless, we can insert them into the answer for the angular speed because we know that the motion is circular. Also note that, if an earth mover with much larger tires, say 1.20 m in radius, were moving at the same speed of 15.0 m/s, its tires would rotate more slowly. They would have an angular speed of

$$\begin{aligned}\omega &= \frac{15.0\ \text{m/s}}{1.20\ \text{m}} \\ &= 12.5\ \text{rad/s}\end{aligned} \tag{6.8}$$

Practice Problems

1. What is the angle in degrees between the hour hand and the minute hand of a clock showing 9:00 a.m.?
   a. 0°
   b. 90°
   c. 180°
   d. 360°

2. What is the approximate value of the arc length between the hour hand and the minute hand of a clock showing 10:00 a.m if the radius of the clock is 0.2 m?
   a. 0.1 m
   b. 0.2 m
   c. 0.3 m
   d. 0.6 m

Check Your Understanding

3. What is circular motion?
   a. Circular motion is the motion of an object when it follows a linear path.
   b. Circular motion is the motion of an object when it follows a zigzag path.
   c. Circular motion is the motion of an object when it follows a circular path.
   d. Circular motion is the movement of an object along the circumference of a circle or rotation along a circular path.

4. What is meant by radius of curvature when describing rotational motion?
   a. The radius of curvature is the radius of a circular path.
   b. The radius of curvature is the diameter of a circular path.
   c. The radius of curvature is the circumference of a circular path.
   d. The radius of curvature is the area of a circular path.

5. What is angular velocity?
   a. Angular velocity is the rate of change of the diameter of the circular path.
   b. Angular velocity is the rate of change of the angle subtended by the circular path.
   c. Angular velocity is the rate of change of the area of the circular path.
   d. Angular velocity is the rate of change of the radius of the circular path.

6. What equation defines angular velocity, ω? Take that r is the radius of curvature, θ is the angle, and t is time.
   a. $\omega = \frac{\Delta\theta}{\Delta t}$
   b. $\omega = \frac{\Delta t}{\Delta\theta}$
   c. $\omega = \frac{\Delta r}{\Delta t}$
   d. $\omega = \frac{\Delta t}{\Delta r}$

7. Identify three examples of an object in circular motion.

a. an artificial satellite orbiting the Earth, a race car moving in the circular race track, and a top spinning on its axis
b. an artificial satellite orbiting the Earth, a race car moving in the circular race track, and a ball tied to a string being swung in a circle around a person's head
c. Earth spinning on its own axis, a race car moving in the circular race track, and a ball tied to a string being swung in a circle around a person's head
d. Earth spinning on its own axis, blades of a working ceiling fan, and a top spinning on its own axis

8. What is the relative orientation of the radius and tangential velocity vectors of an object in uniform circular motion?
a. Tangential velocity vector is always parallel to the radius of the circular path along which the object moves.
b. Tangential velocity vector is always perpendicular to the radius of the circular path along which the object moves.
c. Tangential velocity vector is always at an acute angle to the radius of the circular path along which the object moves.
d. Tangential velocity vector is always at an obtuse angle to the radius of the circular path along which the object moves.

6.2 Uniform Circular Motion

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Describe centripetal acceleration and relate it to linear acceleration
- Describe centripetal force and relate it to linear force
- Solve problems involving centripetal acceleration and centripetal force

Section Key Terms

centrifugal force	centripetal acceleration	centripetal force	uniform circular motion

Centripetal Acceleration

In the previous section, we defined circular motion. The simplest case of circular motion is **uniform circular motion**, where an object travels a circular path *at a constant speed.* Note that, unlike speed, the linear velocity of an object in circular motion is constantly changing because it is always changing direction. We know from kinematics that acceleration is a change in velocity, either in magnitude or in direction or both. Therefore, an object undergoing uniform circular motion is always accelerating, even though the magnitude of its velocity is constant.

You experience this acceleration yourself every time you ride in a car while it turns a corner. If you hold the steering wheel steady during the turn and move at a constant speed, you are executing uniform circular motion. What you notice is a feeling of sliding (or being flung, depending on the speed) away from the center of the turn. This isn't an actual force that is acting on you—it only happens because your body wants to continue moving in a straight line (as per Newton's first law) whereas the car is turning off this straight-line path. Inside the car it appears as if you are forced away from the center of the turn. This fictitious force is known as the **centrifugal force**. The sharper the curve and the greater your speed, the more noticeable this effect becomes.

Figure 6.7 shows an object moving in a circular path at constant speed. The direction of the instantaneous tangential velocity is shown at two points along the path. Acceleration is in the direction of the change in velocity; in this case it points roughly toward the center of rotation. (The center of rotation is at the center of the circular path). If we imagine Δs becoming smaller and smaller, then the acceleration would point *exactly* toward the center of rotation, but this case is hard to draw. We call the acceleration of an object moving in uniform circular motion the **centripetal acceleration** $\mathbf{a}_c$ because centripetal means *center seeking*.

Figure 6.7 The directions of the velocity of an object at two different points are shown, and the change in velocity $\Delta \mathbf{v}$ is seen to point approximately toward the center of curvature (see small inset). For an extremely small value of Δs, $\Delta \mathbf{v}$ points exactly toward the center of the circle (but this is hard to draw). Because $\mathbf{a}_c = \Delta \mathbf{v}/\Delta t$, the acceleration is also toward the center, so $\mathbf{a}_c$ is called centripetal acceleration.

Now that we know that the direction of centripetal acceleration is toward the center of rotation, let's discuss the magnitude of centripetal acceleration. For an object traveling at speed v in a circular path with radius r, the magnitude of centripetal acceleration is

$$\mathbf{a}_c = \frac{v^2}{r}.$$

Centripetal acceleration is greater at high speeds and in sharp curves (smaller radius), as you may have noticed when driving a car, because the car actually pushes you toward the center of the turn. But it is a bit surprising that $\mathbf{a}_c$ is proportional to the speed squared. This means, for example, that the acceleration is four times greater when you take a curve at 100 km/h than at 50 km/h.

We can also express $\mathbf{a}_c$ in terms of the magnitude of angular velocity. Substituting $v = r\omega$ into the equation above, we get $a_c = \frac{(r\omega)^2}{r} = r\omega^2$. Therefore, the magnitude of centripetal acceleration in terms of the magnitude of angular velocity is

$$\mathbf{a}_c = r\omega^2. \tag{6.9}$$

TIPS FOR SUCCESS

The equation expressed in the form $a_c = r\omega^2$ is useful for solving problems where you know the angular velocity rather than the tangential velocity.

Virtual Physics

Ladybug Motion in 2D

In this simulation, you experiment with the position, velocity, and acceleration of a ladybug in circular and elliptical motion. Switch the type of motion from linear to circular and observe the velocity and acceleration vectors. Next, try elliptical motion and notice how the velocity and acceleration vectors differ from those in circular motion.

Click to view content (https://archive.cnx.org/specials/317a2b1e-2fbd-11e5-99b5-e38ffb545fe6/ladybug-motion/)

GRASP CHECK

In uniform circular motion, what is the angle between the acceleration and the velocity? What type of acceleration does a body experience in the uniform circular motion?

a. The angle between acceleration and velocity is 0°, and the body experiences linear acceleration.
b. The angle between acceleration and velocity is 0°, and the body experiences centripetal acceleration.
c. The angle between acceleration and velocity is 90°, and the body experiences linear acceleration.
d. The angle between acceleration and velocity is 90°, and the body experiences centripetal acceleration.

Centripetal Force

Because an object in uniform circular motion undergoes constant acceleration (by changing direction), we know from Newton's second law of motion that there must be a constant net external force acting on the object.

Any force or combination of forces can cause a centripetal acceleration. Just a few examples are the tension in the rope on a tether ball, the force of Earth's gravity on the Moon, the friction between a road and the tires of a car as it goes around a curve, or the normal force of a roller coaster track on the cart during a loop-the-loop.

Any net force causing uniform circular motion is called a **centripetal force**. The direction of a centripetal force is toward the center of rotation, the same as for centripetal acceleration. According to Newton's second law of motion, a net force causes the acceleration of mass according to $\mathbf{F}_{\text{net}} = m\mathbf{a}$. For uniform circular motion, the acceleration is centripetal acceleration: $\mathbf{a} = \mathbf{a}_{\text{c}}$. Therefore, the magnitude of centripetal force, $\mathbf{F}_{\text{c}}$, is $\mathbf{F}_{\text{c}} = m\mathbf{a}_{\text{c}}$.

By using the two different forms of the equation for the magnitude of centripetal acceleration, $\mathbf{a}_{\text{c}} = v^2/r$ and $\mathbf{a}_c = r\omega^2$, we get two expressions involving the magnitude of the centripetal force $\mathbf{F}_{\text{c}}$. The first expression is in terms of tangential speed, the second is in terms of angular speed: $\mathbf{F}_{\text{c}} = m\frac{v^2}{r}$ and $\mathbf{F}_{\text{c}} = mr\omega^2$.

Both forms of the equation depend on mass, velocity, and the radius of the circular path. You may use whichever expression for centripetal force is more convenient. Newton's second law also states that the object will accelerate in the same direction as the net force. By definition, the centripetal force is directed towards the center of rotation, so the object will also accelerate towards the center. A straight line drawn from the circular path to the center of the circle will always be perpendicular to the tangential velocity. Note that, if you solve the first expression for r, you get

$$r = \frac{mv^2}{\mathbf{F}_{\text{c}}}.$$

From this expression, we see that, for a given mass and velocity, a large centripetal force causes a small radius of curvature—that is, a tight curve.

Figure 6.8 In this figure, the frictional force ***f*** serves as the centripetal force $\mathbf{F}_c$. Centripetal force is perpendicular to tangential velocity and causes uniform circular motion. The larger the centripetal force $\mathbf{F}_c$, the smaller is the radius of curvature r and the sharper is the curve. The lower curve has the same velocity **v**, but a larger centripetal force $\mathbf{F}_c$ produces a smaller radius r' .

WATCH PHYSICS

Centripetal Force and Acceleration Intuition

This video explains why a centripetal force creates centripetal acceleration and uniform circular motion. It also covers the difference between speed and velocity and shows examples of uniform circular motion.

Click to view content (https://www.youtube.com/embed/vZOk8NnjILg)

GRASP CHECK

Imagine that you are swinging a yoyo in a vertical clockwise circle in front of you, perpendicular to the direction you are facing. Now, imagine that the string breaks just as the yoyo reaches its bottommost position, nearest the floor. Which of the following describes the path of the yoyo after the string breaks?

a. The yoyo will fly upward in the direction of the centripetal force.
b. The yoyo will fly downward in the direction of the centripetal force.

c. The yoyo will fly to the left in the direction of the tangential velocity.
d. The yoyo will fly to the right in the direction of the tangential velocity.

Solving Centripetal Acceleration and Centripetal Force Problems

To get a feel for the typical magnitudes of centripetal acceleration, we'll do a lab estimating the centripetal acceleration of a tennis racket and then, in our first Worked Example, compare the centripetal acceleration of a car rounding a curve to gravitational acceleration. For the second Worked Example, we'll calculate the force required to make a car round a curve.

Snap Lab

Estimating Centripetal Acceleration

In this activity, you will measure the swing of a golf club or tennis racket to estimate the centripetal acceleration of the end of the club or racket. You may choose to do this in slow motion. Recall that the equation for centripetal acceleration is $\mathbf{a}_c = \frac{v^2}{r}$ or $\mathbf{a}_c = r\omega^2$.

- One tennis racket or golf club
- One timer
- One ruler or tape measure

Procedure

1. Work with a partner. Stand a safe distance away from your partner as he or she swings the golf club or tennis racket.
2. Describe the motion of the swing—is this uniform circular motion? Why or why not?
3. Try to get the swing as close to uniform circular motion as possible. What adjustments did your partner need to make?
4. Measure the radius of curvature. What did you physically measure?
5. By using the timer, find either the linear or angular velocity, depending on which equation you decide to use.
6. What is the approximate centripetal acceleration based on these measurements? How accurate do you think they are? Why? How might you and your partner make these measurements more accurate?

GRASP CHECK

Was it more useful to use the equation $a_c = \frac{v^2}{r}$ or $a_c = r\omega^2$ in this activity? Why?

a. It should be simpler to use $a_c = r\omega^2$ because measuring angular velocity through observation would be easier.
b. It should be simpler to use $a_c = \frac{v^2}{r}$ because measuring tangential velocity through observation would be easier.
c. It should be simpler to use $a_c = r\omega^2$ because measuring angular velocity through observation would be difficult.
d. It should be simpler to use $a_c = \frac{v^2}{r}$ because measuring tangential velocity through observation would be difficult.

WORKED EXAMPLE

Comparing Centripetal Acceleration of a Car Rounding a Curve with Acceleration Due to Gravity

A car follows a curve of radius 500 m at a speed of 25.0 m/s (about 90 km/h). What is the magnitude of the car's centripetal acceleration? Compare the centripetal acceleration for this fairly gentle curve taken at highway speed with acceleration due to gravity (g).

Car around corner

Strategy

Because linear rather than angular speed is given, it is most convenient to use the expression $\mathbf{a}_c = \frac{v^2}{r}$ to find the magnitude of the centripetal acceleration.

Solution

Entering the given values of $v = 25.0$ m/s and $r = 500$ m into the expression for $\mathbf{a}_c$ gives

$$\begin{aligned} \mathbf{a}_c &= \frac{v^2}{r} \\ &= \frac{(25.0\ \text{m/s})^2}{500\ \text{m}} \\ &= 1.25\text{m/s}^2. \end{aligned}$$

Discussion

To compare this with the acceleration due to gravity ($g = 9.80$ m/s^2), we take the ratio $\mathbf{a}_c/g = (1.25\ \text{m/s}^2)/(9.80\text{m/s}^2) = 0.128$. Therefore, $\mathbf{a}_c = 0.128g$, which means that the centripetal acceleration is about one tenth the acceleration due to gravity.

WORKED EXAMPLE

Frictional Force on Car Tires Rounding a Curve

a. Calculate the centripetal force exerted on a 900 kg car that rounds a 600-m-radius curve on horizontal ground at 25.0 m/s.
b. Static friction prevents the car from slipping. Find the magnitude of the frictional force between the tires and the road that allows the car to round the curve without sliding off in a straight line.

Strategy and Solution for (a)

We know that $\mathbf{F}_c = m\frac{v^2}{r}$. Therefore,

$$\begin{aligned}\mathbf{F}_c &= m\frac{v^2}{r} \\ &= \frac{(900\text{ kg})(25.0\text{ m/s})^2}{600\text{ m}} \\ &= 938\text{ N}.\end{aligned}$$

Strategy and Solution for (b)

The image above shows the forces acting on the car while rounding the curve. In this diagram, the car is traveling into the page as shown and is turning to the left. Friction acts toward the left, accelerating the car toward the center of the curve. Because friction is the only horizontal force acting on the car, it provides all of the centripetal force in this case. Therefore, the force of friction is the centripetal force in this situation and points toward the center of the curve.

$$f = \mathbf{F}_c = 938\text{ N}$$

Discussion

Since we found the force of friction in part (b), we could also solve for the coefficient of friction, since $f = \mu_s \mathrm{N} = \mu_s mg$.

Practice Problems

9. What is the centripetal acceleration of an object with speed 12 m/s going along a path of radius 2.0 m?
   a. 6 m/s
   b. 72 m/s
   c. 6 m/s^2
   d. 72 m/s^2

10. Calculate the centripetal acceleration of an object following a path with a radius of a curvature of 0.2 m and at an angular velocity of 5 rad/s.
   a. 1 m/s
   b. 5 m/s
   c. 1 m/s^2
   d. 5 m/s^2

Check Your Understanding

11. What is uniform circular motion?

a. Uniform circular motion is when an object accelerates on a circular path at a constantly increasing velocity.
b. Uniform circular motion is when an object travels on a circular path at a variable acceleration.
c. Uniform circular motion is when an object travels on a circular path at a constant speed.
d. Uniform circular motion is when an object travels on a circular path at a variable speed.

12. What is centripetal acceleration?
a. The acceleration of an object moving in a circular path and directed radially toward the center of the circular orbit
b. The acceleration of an object moving in a circular path and directed tangentially along the circular path
c. The acceleration of an object moving in a linear path and directed in the direction of motion of the object
d. The acceleration of an object moving in a linear path and directed in the direction opposite to the motion of the object

13. Is there a net force acting on an object in uniform circular motion?
a. Yes, the object is accelerating, so a net force must be acting on it.
b. Yes, because there is no acceleration.
c. No, because there is acceleration.
d. No, because there is no acceleration.

14. Identify two examples of forces that can cause centripetal acceleration.
a. The force of Earth's gravity on the moon and the normal force
b. The force of Earth's gravity on the moon and the tension in the rope on an orbiting tetherball
c. The normal force and the force of friction acting on a moving car
d. The normal force and the tension in the rope on a tetherball

6.3 Rotational Motion

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Describe rotational kinematic variables and equations and relate them to their linear counterparts
- Describe torque and lever arm
- Solve problems involving torque and rotational kinematics

Section Key Terms

angular acceleration	kinematics of rotational motion	lever arm
tangential acceleration	torque	

Rotational Kinematics

In the section on uniform circular motion, we discussed motion in a circle at constant speed and, therefore, constant angular velocity. However, there are times when angular velocity is not constant—rotational motion can speed up, slow down, or reverse directions. Angular velocity is not constant when a spinning skater pulls in her arms, when a child pushes a merry-go-round to make it rotate, or when a CD slows to a halt when switched off. In all these cases, **angular acceleration** occurs because the angular velocity $\boldsymbol{\omega}$ changes. The faster the change occurs, the greater is the angular acceleration. Angular acceleration $\boldsymbol{\alpha}$ is the rate of change of angular velocity. In equation form, angular acceleration is

$$\boldsymbol{\alpha} = \frac{\Delta\boldsymbol{\omega}}{\Delta t},$$

where $\Delta\boldsymbol{\omega}$ is the change in angular velocity and Δt is the change in time. The units of angular acceleration are (rad/s)/s, or rad/s^2. If $\boldsymbol{\omega}$ increases, then $\boldsymbol{\alpha}$ is positive. If $\boldsymbol{\omega}$ decreases, then $\boldsymbol{\alpha}$ is negative. Keep in mind that, by convention, counterclockwise is the positive direction and clockwise is the negative direction. For example, the skater in Figure 6.9 is rotating counterclockwise as seen from above, so her angular velocity is positive. Acceleration would be negative, for example, when an object that is rotating counterclockwise slows down. It would be positive when an object that is rotating counterclockwise speeds up.

Figure 6.9 A figure skater spins in the counterclockwise direction, so her angular velocity is normally considered to be positive. (Luu, Wikimedia Commons)

The relationship between the magnitudes of **tangential acceleration**, $\mathbf{a}$, and angular acceleration,

$$\boldsymbol{\alpha}, \text{is} \mathbf{a} = r\boldsymbol{\alpha} \text{or} \boldsymbol{\alpha} = \frac{\mathbf{a}}{r}. \tag{6.10}$$

These equations mean that the magnitudes of tangential acceleration and angular acceleration are directly proportional to each other. The greater the angular acceleration, the larger the change in tangential acceleration, and vice versa. For example, consider riders in their pods on a Ferris wheel at rest. A Ferris wheel with greater angular acceleration will give the riders greater tangential acceleration because, as the Ferris wheel increases its rate of spinning, it also increases its tangential velocity. Note that the radius of the spinning object also matters. For example, for a given angular acceleration $\boldsymbol{\alpha}$, a smaller Ferris wheel leads to a smaller tangential acceleration for the riders.

TIPS FOR SUCCESS

Tangential acceleration is sometimes denoted $\mathbf{a}_t$. It is a linear acceleration in a direction tangent to the circle at the point of interest in circular or rotational motion. Remember that tangential acceleration is parallel to the tangential velocity (either in the same direction or in the opposite direction.) Centripetal acceleration is always perpendicular to the tangential velocity.

So far, we have defined three rotational variables: θ, $\boldsymbol{\omega}$, and $\boldsymbol{\alpha}$. These are the angular versions of the linear variables x, $\mathbf{v}$, and $\mathbf{a}$. Table 6.2 shows how they are related.

Rotational	Linear	Relationship
θ	x	$\theta = \frac{x}{r}$

Table 6.2 Rotational and Linear Variables

Rotational	Linear	Relationship
$\boldsymbol{\omega}$	$\mathbf{v}$	$\boldsymbol{\omega} = \frac{\mathbf{v}}{r}$
$\boldsymbol{\alpha}$	$\mathbf{a}$	$\boldsymbol{\alpha} = \frac{\mathbf{a}}{r}$

Table 6.2 Rotational and Linear Variables

We can now begin to see how rotational quantities like θ, $\boldsymbol{\omega}$, and $\boldsymbol{\alpha}$ are related to each other. For example, if a motorcycle wheel that starts at rest has a large angular acceleration for a fairly long time, it ends up spinning rapidly and rotates through many revolutions. Putting this in terms of the variables, if the wheel's angular acceleration $\boldsymbol{\alpha}$ is large for a long period of time t, then the final angular velocity $\boldsymbol{\omega}$ and angle of rotation θ are large. In the case of linear motion, if an object starts at rest and undergoes a large linear acceleration, then it has a large final velocity and will have traveled a large distance.

The **kinematics of rotational motion** describes the relationships between the angle of rotation, angular velocity, angular acceleration, and time. It only *describes* motion—it does not include any forces or masses that may affect rotation (these are part of dynamics). Recall the kinematics equation for linear motion: $\mathbf{v} = \mathbf{v}_0 + \mathbf{a}t$ (constant $\mathbf{a}$).

As in linear kinematics, we assume $\mathbf{a}$ is constant, which means that angular acceleration $\boldsymbol{\alpha}$ is also a constant, because $\mathbf{a} = r\boldsymbol{\alpha}$. The equation for the kinematics relationship between $\boldsymbol{\omega}$, $\boldsymbol{\alpha}$, and t is

$$\boldsymbol{\omega} = \boldsymbol{\omega}_0 + \boldsymbol{\alpha}t(\text{constant}\alpha),$$

where $\boldsymbol{\omega}_0$ is the initial angular velocity. Notice that the equation is identical to the linear version, except with angular analogs of the linear variables. In fact, all of the linear kinematics equations have rotational analogs, which are given in Table 6.3. These equations can be used to solve rotational or linear kinematics problem in which $\mathbf{a}$ and $\boldsymbol{\alpha}$ are constant.

Rotational	Linear	
$\theta = \overline{\boldsymbol{\omega}}t$	$x = \overline{\mathbf{v}}t$	
$\boldsymbol{\omega} = \boldsymbol{\omega}_0 + \boldsymbol{\alpha}t$	$\mathbf{v} = \mathbf{v}_0 + \boldsymbol{\alpha}t$	constant $\boldsymbol{\alpha}$, $\mathbf{a}$
$\theta = \boldsymbol{\omega}_0 t + \frac{1}{2}\boldsymbol{\alpha}t^2$	$x = \mathbf{v}_0 t + \frac{1}{2}\boldsymbol{\alpha}t^2$	constant $\boldsymbol{\alpha}$, $\mathbf{a}$
$\boldsymbol{\omega}^2 = \boldsymbol{\omega}_0{}^2 + 2\boldsymbol{\alpha}\theta$	$\mathbf{v}^2 = \mathbf{v}_0{}^2 + 2\boldsymbol{\alpha}x$	constant $\boldsymbol{\alpha}$, $\mathbf{a}$

Table 6.3 Equations for Rotational Kinematics

In these equations, $\boldsymbol{\omega}_0$ and $\mathbf{v}_0$ are initial values, t_0 is zero, and the average angular velocity $\overline{\boldsymbol{\omega}}$ and average velocity $\overline{\mathbf{v}}$ are

$$\overline{\boldsymbol{\omega}} = \frac{\boldsymbol{\omega}_0 + \boldsymbol{\omega}}{2} \text{and} \overline{\mathbf{v}} = \frac{\mathbf{v}_0 + \mathbf{v}}{2}. \tag{6.11}$$

FUN IN PHYSICS

Storm Chasing

Figure 6.10 Tornadoes descend from clouds in funnel-like shapes that spin violently. (Daphne Zaras, U.S. National Oceanic and Atmospheric Administration)

Storm chasers tend to fall into one of three groups: Amateurs chasing tornadoes as a hobby, atmospheric scientists gathering data for research, weather watchers for news media, or scientists having fun under the guise of work. Storm chasing is a dangerous pastime because tornadoes can change course rapidly with little warning. Since storm chasers follow in the wake of the destruction left by tornadoes, changing flat tires due to debris left on the highway is common. The most active part of the world for tornadoes, called *tornado alley*, is in the central United States, between the Rocky Mountains and Appalachian Mountains.

Tornadoes are perfect examples of rotational motion in action in nature. They come out of severe thunderstorms called supercells, which have a column of air rotating around a horizontal axis, usually about four miles across. The difference in wind speeds between the strong cold winds higher up in the atmosphere in the jet stream and weaker winds traveling north from the Gulf of Mexico causes the column of rotating air to shift so that it spins around a vertical axis, creating a tornado.

Tornadoes produce wind speeds as high as 500 km/h (approximately 300 miles/h), particularly at the bottom where the funnel is narrowest because the rate of rotation increases as the radius decreases. They blow houses away as if they were made of paper and have been known to pierce tree trunks with pieces of straw.

GRASP CHECK

What is the physics term for the eye of the storm? Why would winds be weaker at the eye of the tornado than at its outermost edge?

a. The eye of the storm is the center of rotation. Winds are weaker at the eye of a tornado because tangential velocity is directly proportional to radius of curvature.
b. The eye of the storm is the center of rotation. Winds are weaker at the eye of a tornado because tangential velocity is inversely proportional to radius of curvature.
c. The eye of the storm is the center of rotation. Winds are weaker at the eye of a tornado because tangential velocity is directly proportional to the square of the radius of curvature.
d. The eye of the storm is the center of rotation. Winds are weaker at the eye of a tornado because tangential velocity is inversely proportional to the square of the radius of curvature.

Torque

If you have ever spun a bike wheel or pushed a merry-go-round, you know that force is needed to change angular velocity. The farther the force is applied from the pivot point (or fulcrum), the greater the angular acceleration. For example, a door opens slowly if you push too close to its hinge, but opens easily if you push far from the hinges. Furthermore, we know that the more

massive the door is, the more slowly it opens; this is because angular acceleration is inversely proportional to mass. These relationships are very similar to the relationships between force, mass, and acceleration from Newton's second law of motion. Since we have already covered the angular versions of distance, velocity and time, you may wonder what the angular version of force is, and how it relates to linear force.

The angular version of force is **torque** $\boldsymbol{\tau}$, which is the turning effectiveness of a force. See Figure 6.11. The equation for the magnitude of torque is

$$\boldsymbol{\tau} = r\mathbf{F}\sin\theta,$$

where r is the magnitude of the **lever arm**, $\mathbf{F}$ is the magnitude of the linear force, and θ is the angle between the lever arm and the force. The lever arm is the vector from the point of rotation (pivot point or fulcrum) to the location where force is applied. Since the magnitude of the lever arm is a distance, its units are in meters, and torque has units of N·m. Torque is a vector quantity and has the same direction as the angular acceleration that it produces.

Figure 6.11 A man pushes a merry-go-round at its edge and perpendicular to the lever arm to achieve maximum torque.

Applying a stronger torque will produce a greater angular acceleration. For example, the harder the man pushes the merry-go-round in Figure 6.11, the faster it accelerates. Furthermore, the more massive the merry-go-round is, the slower it accelerates for the same torque. If the man wants to maximize the effect of his force on the merry-go-round, he should push as far from the center as possible to get the largest lever arm and, therefore, the greatest torque and angular acceleration. Torque is also maximized when the force is applied perpendicular to the lever arm.

Solving Rotational Kinematics and Torque Problems

Just as linear forces can balance to produce zero net force and no linear acceleration, the same is true of rotational motion. When two torques of equal magnitude act in opposing directions, there is no net torque and no angular acceleration, as you can see in the following video. If zero net torque acts on a system spinning at a constant angular velocity, the system will continue to spin at the same angular velocity.

WATCH PHYSICS

Introduction to Torque

This video (https://www.khanacademy.org/science/physics/torque-angular-momentum/torque-tutorial/v/introduction-to-torque) defines torque in terms of moment arm (which is the same as lever arm). It also covers a problem with forces acting in opposing directions about a pivot point. (At this stage, you can ignore Sal's references to work and mechanical advantage.)

GRASP CHECK

Click to view content (https://www.openstax.org/l/28torque)

If the net torque acting on the ruler from the example was positive instead of zero, what would this say about the angular

acceleration? What would happen to the ruler over time?

a. The ruler is in a state of rotational equilibrium so it will not rotate about its center of mass. Thus, the angular acceleration will be zero.
b. The ruler is not in a state of rotational equilibrium so it will not rotate about its center of mass. Thus, the angular acceleration will be zero.
c. The ruler is not in a state of rotational equilibrium so it will rotate about its center of mass. Thus, the angular acceleration will be non-zero.
d. The ruler is in a state of rotational equilibrium so it will rotate about its center of mass. Thus, the angular acceleration will be non-zero.

Now let's look at examples applying rotational kinematics to a fishing reel and the concept of torque to a merry-go-round.

WORKED EXAMPLE

Calculating the Time for a Fishing Reel to Stop Spinning

A deep-sea fisherman uses a fishing rod with a reel of radius 4.50 cm. A big fish takes the bait and swims away from the boat, pulling the fishing line from his fishing reel. As the fishing line unwinds from the reel, the reel spins at an angular velocity of 220 rad/s. The fisherman applies a brake to the spinning reel, creating an angular acceleration of –300 rad/s². How long does it take the reel to come to a stop?

Strategy

We are asked to find the time *t* for the reel to come to a stop. The magnitude of the initial angular velocity is $\boldsymbol{\omega}_0 = 220$ rad/s, and the magnitude of the final angular velocity $\boldsymbol{\omega} = 0$. The signed magnitude of the angular acceleration is $\boldsymbol{\alpha} = -300$ rad/s², where the minus sign indicates that it acts in the direction opposite to the angular velocity. Looking at the rotational kinematic equations, we see all quantities but *t* are known in the equation $\boldsymbol{\omega} = \boldsymbol{\omega}_0 + \boldsymbol{\alpha} t$, making it the easiest equation to use for this problem.

Solution

The equation to use is $\boldsymbol{\omega} = \boldsymbol{\omega}_0 + \boldsymbol{\alpha} t$.

We solve the equation algebraically for *t*, and then insert the known values.

$$\begin{aligned} t &= \frac{\omega - \omega_0}{\alpha} \\ &= \frac{0 - 220 \text{ rad/s}}{-300 \text{rad/s}^2} \\ &= 0.733 \text{ s} \end{aligned} \tag{6.12}$$

Discussion

The time to stop the reel is fairly small because the acceleration is fairly large. Fishing lines sometimes snap because of the forces involved, and fishermen often let the fish swim for a while before applying brakes on the reel. A tired fish will be slower, requiring a smaller acceleration and therefore a smaller force.

WORKED EXAMPLE

Calculating the Torque on a Merry-Go-Round

Consider the man pushing the playground merry-go-round in Figure 6.11. He exerts a force of 250 N at the edge of the merry-go-round and perpendicular to the radius, which is 1.50 m. How much torque does he produce? Assume that friction acting on the merry-go-round is negligible.

Strategy

To find the torque, note that the applied force is perpendicular to the radius and that friction is negligible.

Solution

$$\begin{aligned} \tau &= r\mathbf{F}\sin\theta \\ &= (1.50\text{ m})(250\text{ N})\sin\left(\tfrac{\pi}{2}\right). \\ &= 375\text{ N}\cdot\text{m} \end{aligned} \tag{6.13}$$

Discussion

The man maximizes the torque by applying force perpendicular to the lever arm, so that $\theta = \frac{\pi}{2}$ and $\sin\theta = 1$. The man also maximizes his torque by pushing at the outer edge of the merry-go-round, so that he gets the largest-possible lever arm.

Practice Problems

15. How much torque does a person produce if he applies a 12 N force 1.0 m away from the pivot point, perpendicularly to the lever arm?
    a. $\frac{1}{144}$ N-m
    b. $\frac{1}{12}$ N-m
    c. 12 N-m
    d. 144 N-m

16. An object's angular velocity changes from 3 rad/s clockwise to 8 rad/s clockwise in 5 s. What is its angular acceleration?
    a. 0.6 rad/s^2
    b. 1.6 rad/s^2
    c. 1 rad/s^2
    d. 5 rad/s^2

Check Your Understanding

17. What is angular acceleration?
    a. Angular acceleration is the rate of change of the angular displacement.
    b. Angular acceleration is the rate of change of the angular velocity.
    c. Angular acceleration is the rate of change of the linear displacement.
    d. Angular acceleration is the rate of change of the linear velocity.

18. What is the equation for angular acceleration, α? Assume θ is the angle, ω is the angular velocity, and t is time.
    a. $\alpha = \frac{\Delta\omega}{\Delta t}$
    b. $\alpha = \Delta\omega\Delta t$
    c. $\alpha = \frac{\Delta\theta}{\Delta t}$
    d. $\alpha = \Delta\theta\Delta t$

19. Which of the following best describes torque?
    a. It is the rotational equivalent of a force.
    b. It is the force that affects linear motion.
    c. It is the rotational equivalent of acceleration.
    d. It is the acceleration that affects linear motion.

20. What is the equation for torque?

a. $\tau = F\,cos\theta\,r$
b. $\tau = \frac{F\sin\theta}{r}$
c. $\tau = rF\cos\theta$
d. $\tau = rF\sin\theta$

KEY TERMS

angle of rotation the ratio of the arc length to the radius of curvature of a circular path

angular acceleration the rate of change of angular velocity with time

angular velocity ($\boldsymbol{\omega}$) the rate of change in the angular position of an object following a circular path

arc length (Δs) the distance traveled by an object along a circular path

centrifugal force a fictitious force that acts in the direction opposite the centripetal acceleration

centripetal acceleration the acceleration of an object moving in a circle, directed toward the center of the circle

centripetal force any force causing uniform circular motion

circular motion the motion of an object along a circular path

kinematics of rotational motion the relationships between rotation angle, angular velocity, angular acceleration, and time

lever arm the distance between the point of rotation (pivot point) and the location where force is applied

radius of curvature the distance between the center of a circular path and the path

rotational motion the circular motion of an object about an axis of rotation

spin rotation about an axis that goes through the center of mass of the object

tangential acceleration the acceleration in a direction tangent to the circular path of motion and in the same direction or opposite direction as the tangential velocity

tangential velocity the instantaneous linear velocity of an object in circular or rotational motion

torque the effectiveness of a force to change the rotational speed of an object

uniform circular motion the motion of an object in a circular path at constant speed

SECTION SUMMARY

6.1 Angle of Rotation and Angular Velocity

- Circular motion is motion in a circular path.
- The angle of rotation $\Delta\theta$ is defined as the ratio of the arc length to the radius of curvature.
- The arc length Δs is the distance traveled along a circular path and r is the radius of curvature of the circular path.
- The angle of rotation $\Delta\theta$ is measured in units of radians (rad), where $2\ \pi\text{rad} = 360° = 1$ revolution.
- Angular velocity $\boldsymbol{\omega}$ is the rate of change of an angle, where a rotation $\Delta\theta$ occurs in a time Δt .
- The units of angular velocity are radians per second (rad/s).
- Tangential speed v and angular speed ω are related by $v = r\omega$, and tangential velocity has units of m/s.
- The direction of angular velocity is along the axis of rotation, toward (away) from you for clockwise (counterclockwise) motion.

6.2 Uniform Circular Motion

- Centripetal acceleration $\mathbf{a}_c$ is the acceleration experienced while in uniform circular motion.
- Centripetal acceleration force is a *center-seeking* force that always points toward the center of rotation, perpendicular to the linear velocity, in the same direction as the net force, and in the direction opposite that of the radius vector.
- The standard unit for centripetal acceleration is m/s^2.
- Centripetal force $\mathbf{F}_c$ is any net force causing uniform circular motion.

6.3 Rotational Motion

- Kinematics is the description of motion.
- The kinematics of rotational motion describes the relationships between rotation angle, angular velocity, angular acceleration, and time.
- Torque is the effectiveness of a force to change the rotational speed of an object. Torque is the rotational analog of force.
- The lever arm is the distance between the point of rotation (pivot point) and the location where force is applied.
- Torque is maximized by applying force perpendicular to the lever arm and at a point as far as possible from the pivot point or fulcrum. If torque is zero, angular acceleration is zero.

KEY EQUATIONS

6.1 Angle of Rotation and Angular Velocity

Angle of rotation	$\Delta\theta = \frac{\Delta s}{r}$
Angular speed:	$\omega = \frac{\Delta\theta}{\Delta t}$
Tangential speed:	$v = r\omega$

6.2 Uniform Circular Motion

Centripetal acceleration	$\mathbf{a}_c = \frac{v^2}{r}$ or $\mathbf{a}_c = r\omega^2$
Centripetal force	$\mathbf{F}_c = m\mathbf{a}_c, \mathbf{F}_c = m\frac{v^2}{r}, \mathbf{F}_c = mr\omega^2$

6.3 Rotational Motion

Angular acceleration	$\boldsymbol{\alpha} = \frac{\Delta\boldsymbol{\omega}}{\Delta t}$
Rotational kinematic equations	$\theta = \boldsymbol{\omega}t, \boldsymbol{\omega} = \boldsymbol{\omega}_0 + \boldsymbol{\alpha}t, \theta = \boldsymbol{\omega}_0 t + \frac{1}{2}\boldsymbol{\alpha}t^2, \boldsymbol{\omega}^2 = \boldsymbol{\omega}_0{}^2 + 2\boldsymbol{\alpha}\theta$
Tangential (linear) acceleration	$\mathbf{a} = r\boldsymbol{\alpha}$
Torque	$\boldsymbol{\tau} = r\mathbf{F}\sin\theta$

CHAPTER REVIEW

Concept Items

6.1 Angle of Rotation and Angular Velocity

1. One revolution is equal to how many radians? Degrees?
   a. $1 \text{ rev} = \pi \text{ rad} = 180°$
   b. $1 \text{ rev} = \pi \text{ rad} = 360°$
   c. $1 \text{ rev} = 2\pi \text{ rad} = 180°$
   d. $1 \text{ rev} = 2\pi \text{ rad} = 360°$

2. What is tangential velocity?
   a. Tangential velocity is the average linear velocity of an object in a circular motion.
   b. Tangential velocity is the instantaneous linear velocity of an object undergoing rotational motion.
   c. Tangential velocity is the average angular velocity of an object in a circular motion.
   d. Tangential velocity is the instantaneous angular velocity of an object in a circular motion.

3. What kind of motion is called *spin*?
   a. Spin is rotational motion of an object about an axis parallel to the axis of the object.
   b. Spin is translational motion of an object about an axis parallel to the axis of the object.
   c. Spin is the rotational motion of an object about its center of mass.
   d. Spin is translational motion of an object about its own axis.

6.2 Uniform Circular Motion

4. What is the equation for centripetal acceleration in terms of angular velocity and the radius?
   a. $a_c = \frac{\omega^2}{r}$
   b. $a_c = \frac{\omega}{r}$
   c. $a_c = r\omega^2$
   d. $a_c = r\omega$

5. How can you express centripetal force in terms of centripetal acceleration?
   a. $F_c = \frac{a_c^2}{m}$
   b. $F_c = \frac{a_c}{m}$
   c. $F_c = ma_c^2$
   d. $F_c = ma_c$

6. What is meant by the word centripetal?
   a. center-seeking
   b. center-avoiding
   c. central force
   d. central acceleration

6.3 Rotational Motion

7. Conventionally, for which direction of rotation of an object is angular acceleration considered positive?
   a. the positive *x* direction of the coordinate system
   b. the negative *x* direction of the coordinate system
   c. the counterclockwise direction
   d. the clockwise direction

8. When you push a door closer to the hinges, why does it open more slowly?
   a. It opens slowly, because the lever arm is shorter so the torque is large.
   b. It opens slowly because the lever arm is longer so the torque is large.
   c. It opens slowly, because the lever arm is shorter so the torque is less.
   d. It opens slowly, because the lever arm is longer so the torque is less.

9. When is angular acceleration negative?
   a. Angular acceleration is the rate of change of the displacement and is negative when ω increases.
   b. Angular acceleration is the rate of change of the displacement and is negative when ω decreases.
   c. Angular acceleration is the rate of change of angular velocity and is negative when ω increases.
   d. Angular acceleration is the rate of change of angular velocity and is negative when ω decreases.

Critical Thinking Items

6.1 Angle of Rotation and Angular Velocity

10. When the radius of the circular path of rotational motion increases, what happens to the arc length for a given angle of rotation?
   a. The arc length is directly proportional to the radius of the circular path, and it increases with the radius.
   b. The arc length is inversely proportional to the radius of the circular path, and it decreases with the radius.
   c. The arc length is directly proportional to the radius of the circular path, and it decreases with the radius.
   d. The arc length is inversely proportional to the radius of the circular path, and it increases with the radius.

11. Consider a CD spinning clockwise. What is the sum of the instantaneous velocities of two points on both ends of its diameter?
   a. $2v$
   b. $\frac{v}{2}$
   c. $-v$
   d. 0

6.2 Uniform Circular Motion

12. What are the directions of the velocity and acceleration of an object in uniform circular motion?
   a. Velocity is tangential, and acceleration is radially outward.
   b. Velocity is tangential, and acceleration is radially inward.
   c. Velocity is radially outward, and acceleration is tangential.
   d. Velocity is radially inward, and acceleration is tangential.

13. Suppose you have an object tied to a rope and are rotating it over your head in uniform circular motion. If you increase the length of the rope, would you have to apply more or less force to maintain the same speed?
   a. More force is required, because the force is inversely proportional to the radius of the circular orbit.
   b. More force is required because the force is directly proportional to the radius of the circular orbit.
   c. Less force is required because the force is inversely proportional to the radius of the circular orbit.
   d. Less force is required because the force is directly proportional to the radius of the circular orbit.

6.3 Rotational Motion

14. Consider two spinning tops with different radii. Both have the same linear instantaneous velocities at their edges. Which top has a higher angular velocity?
   a. the top with the smaller radius because the radius of curvature is inversely proportional to the angular velocity
   b. the top with the smaller radius because the radius of curvature is directly proportional to the angular velocity
   c. the top with the larger radius because the radius of curvature is inversely proportional to the angular velocity
   d. The top with the larger radius because the radius of curvature is directly proportional to the angular velocity

15. A person tries to lift a stone by using a lever. If the lever arm is constant and the mass of the stone increases, what is true of the torque necessary to lift it?
   a. It increases, because the torque is directly proportional to the mass of the body.
   b. It increases because the torque is inversely proportional to the mass of the body.
   c. It decreases because the torque is directly proportional to the mass of the body.
   d. It decreases, because the torque is inversely proportional to the mass of the body.

Problems

6.1 Angle of Rotation and Angular Velocity

16. What is the angle of rotation (in degrees) between two hands of a clock, if the radius of the clock is $0.70\ \text{m}$ and the arc length separating the two hands is $1.0\ \text{m}$?
a. 40°
b. 80°
c. 81°
d. 163°

17. A clock has radius of $0.5\ \text{m}$. The outermost point on its minute hand travels along the edge. What is its tangential speed?
a. 9×10^{-4} m/s
b. 3.4×10^{-3} m/s
c. 8.5×10^{-4} m/s
d. 1.3×10^{1} m/s

6.2 Uniform Circular Motion

18. What is the centripetal force exerted on a 1,600 kg car that rounds a 100 m radius curve at 12 m/s?
a. 192 N
b. 1, 111 N
c. 2, 300 N
d. 13, 333 N

19. Find the frictional force between the tires and the road that allows a 1,000 kg car traveling at 30 m/s to round a 20 m radius curve.
a. 22 N
b. 667 N
c. 1, 500 N
d. 45, 000 N

6.3 Rotational Motion

20. An object's angular acceleration is 36 rad/s^2. If it were initially spinning with a velocity of 6.0 m/s, what would its angular velocity be after 5.0 s?
a. 186 rad/s
b. 190 rad/s^2
c. −174 rad/s
d. −174 rad/s^2

21. When a fan is switched on, it undergoes an angular acceleration of 150 rad/s^2. How long will it take to achieve its maximum angular velocity of 50 rad/s?
a. −0.3 s
b. 0.3 s
c. 3.0 s

Performance Task

6.3 Rotational Motion

22. Design a lever arm capable of lifting a 0.5 kg object such as a stone. The force for lifting should be provided by placing coins on the other end of the lever. How many coins would you need? What happens if you shorten or lengthen the lever arm? What does this say about torque?

TEST PREP

Multiple Choice

6.1 Angle of Rotation and Angular Velocity

23. What is 1 radian approximately in degrees?
a. 57.3°
b. 360°
c. π°
d. 2π°

24. If the following objects are spinning at the same angular velocities, the edge of which one would have the highest speed?
a. Mini CD
b. Regular CD
c. Vinyl record

25. What are possible units for tangential velocity?
a. m/s
b. rad/s
c. °/s

26. What is 30° in radians?
a. $\frac{\pi}{12}$
b. $\frac{\pi}{9}$
c. $\frac{\pi}{6}$
d. $\frac{\pi}{3}$

27. For a given object, what happens to the arc length as the angle of rotation increases?
a. The arc length is directly proportional to the angle of rotation, so it increases with the angle of rotation.
b. The arc length is inversely proportional to the angle of rotation, so it decreases with the angle of rotation.
c. The arc length is directly proportional to the angle of rotation, so it decreases with the angle of rotation.

d. The arc length is inversely proportional to the angle of rotation, so it increases with the angle of rotation.

6.2 Uniform Circular Motion

28. Which of these quantities is constant in uniform circular motion?
a. Speed
b. Velocity
c. Acceleration
d. Displacement

29. Which of these quantities impact centripetal force?
a. Mass and speed only
b. Mass and radius only
c. Speed and radius only
d. Mass, speed, and radius all impact centripetal force

30. An increase in the magnitude of which of these quantities causes a reduction in centripetal force?
a. Mass
b. Radius of curvature
c. Speed

31. What happens to centripetal acceleration as the radius of curvature decreases and the speed is constant, and why?
a. It increases, because the centripetal acceleration is inversely proportional to the radius of the curvature.
b. It increases, because the centripetal acceleration is directly proportional to the radius of curvature.
c. It decreases, because the centripetal acceleration is inversely proportional to the radius of the curvature.
d. It decreases, because the centripetal acceleration is directly proportional to the radius of the curvature.

32. Why do we experience more sideways acceleration while driving around sharper curves?
a. Centripetal acceleration is inversely proportional to the radius of curvature, so it increases as the radius of curvature decreases.
b. Centripetal acceleration is directly proportional to the radius of curvature, so it decreases as the radius of curvature decreases.
c. Centripetal acceleration is directly proportional to the radius of curvature, so it decreases as the radius of curvature increases.
d. Centripetal acceleration is directly proportional to the radius of curvature, so it increases as the radius of curvature increases.

6.3 Rotational Motion

33. Which of these quantities is not described by the kinematics of rotational motion?
a. Rotation angle
b. Angular acceleration
c. Centripetal force
d. Angular velocity

34. In the equation $\tau = rF\sin\theta$, what is F?
a. Linear force
b. Centripetal force
c. Angular force

35. What happens when two torques act equally in opposite directions?
a. Angular velocity is zero.
b. Angular acceleration is zero.

36. What is the mathematical relationship between angular and linear accelerations?
a. $a = r\alpha$
b. $a = \frac{\alpha}{r}$
c. $a = r^2\alpha$
d. $a = \frac{\alpha}{r^2}$

Short Answer

6.1 Angle of Rotation and Angular Velocity

37. What is the rotational analog of linear velocity?
a. Angular displacement
b. Angular velocity
c. Angular acceleration
d. Angular momentum

38. What is the rotational analog of distance?
a. Rotational angle
b. Torque
c. Angular velocity
d. Angular momentum

39. What is the equation that relates the linear speed of a point on a rotating object with the object's angular quantities?
a. $v = \frac{\omega}{r}$
b. $v = r\omega$
c. $v = \frac{\alpha}{r}$
d. $v = r\alpha$

40. As the angular velocity of an object increases, what happens to the linear velocity of a point on that object?
a. It increases, because linear velocity is directly proportional to angular velocity.
b. It increases, because linear velocity is inversely proportional to angular velocity.

c. It decreases because linear velocity is directly proportional to angular velocity.
d. It decreases because linear velocity is inversely proportional to angular velocity.

41. What is angular speed in terms of tangential speed and the radius?
a. $\omega = \frac{v^2}{r}$
b. $\omega = \frac{v}{r}$
c. $\omega = rv$
d. $\omega = rv^2$

42. Why are radians dimensionless?
a. Radians are dimensionless, because they are defined as a ratio of distances. They are defined as the ratio of the arc length to the radius of the circle.
b. Radians are dimensionless because they are defined as a ratio of distances. They are defined as the ratio of the area to the radius of the circle.
c. Radians are dimensionless because they are defined as multiplication of distance. They are defined as the multiplication of the arc length to the radius of the circle.
d. Radians are dimensionless because they are defined as multiplication of distance. They are defined as the multiplication of the area to the radius of the circle.

6.2 Uniform Circular Motion

43. What type of quantity is centripetal acceleration?
a. Scalar quantity; centripetal acceleration has magnitude only but no direction
b. Scalar quantity; centripetal acceleration has magnitude as well as direction
c. Vector quantity; centripetal acceleration has magnitude only but no direction
d. Vector quantity; centripetal acceleration has magnitude as well as direction

44. What are the standard units for centripetal acceleration?
a. m/s
b. m/s^2
c. m^2/s
d. m^2/s^2

45. What is the angle formed between the vectors of tangential velocity and centripetal force?
a. 0°
b. 30°
c. 90°
d. 180°

46. What is the angle formed between the vectors of centripetal acceleration and centripetal force?
a. 0°
b. 30°
c. 90°
d. 180°

47. What are the standard units for centripetal force?
a. m
b. m/s
c. m/s^2
d. newtons

48. As the mass of an object in uniform circular motion increases, what happens to the centripetal force required to keep it moving at the same speed?
a. It increases, because the centripetal force is directly proportional to the mass of the rotating body.
b. It increases, because the centripetal force is inversely proportional to the mass of the rotating body.
c. It decreases, because the centripetal force is directly proportional to the mass of the rotating body.
d. It decreases, because the centripetal force is inversely proportional to the mass of the rotating body.

6.3 Rotational Motion

49. The relationships between which variables are described by the kinematics of rotational motion?
a. The kinematics of rotational motion describes the relationships between rotation angle, angular velocity, and angular acceleration.
b. The kinematics of rotational motion describes the relationships between rotation angle, angular velocity, angular acceleration, and angular momentum.
c. The kinematics of rotational motion describes the relationships between rotation angle, angular velocity, angular acceleration, and time.
d. The kinematics of rotational motion describes the relationships between rotation angle, angular velocity, angular acceleration, torque, and time.

50. What is the kinematics relationship between ω, α, and t ?
a. $\omega = \alpha t$
b. $\omega = \omega_0 - \alpha t$
c. $\omega = \omega_0 + \alpha t$
d. $\omega = \omega_0 + \frac{1}{2}\alpha t$

51. What kind of quantity is torque?
a. Scalar
b. Vector

c. Dimensionless
d. Fundamental quantity

52. If a linear force is applied to a lever arm farther away from the pivot point, what happens to the resultant torque?
a. It decreases.
b. It increases.
c. It remains the same.
d. It changes the direction.

53. How can the same force applied to a lever produce different torques?
a. By applying the force at different points of the lever arm along the length of the lever or by changing the angle between the lever arm and the applied force.
b. By applying the force at the same point of the lever arm along the length of the lever or by changing the angle between the lever arm and the applied force.
c. By applying the force at different points of the lever arm along the length of the lever or by maintaining the same angle between the lever arm and the applied force.
d. By applying the force at the same point of the lever arm along the length of the lever or by maintaining the same angle between the lever arm and the applied force.

Extended Response

6.1 Angle of Rotation and Angular Velocity

54. Consider two pits on a CD, one close to the center and one close to the outer edge. When the CD makes one full rotation, which pit would have gone through a greater angle of rotation? Which one would have covered a greater arc length?
a. The one close to the center would go through the greater angle of rotation. The one near the outer edge would trace a greater arc length.
b. The one close to the center would go through the greater angle of rotation. The one near the center would trace a greater arc length.
c. Both would go through the same angle of rotation. The one near the outer edge would trace a greater arc length.
d. Both would go through the same angle of rotation. The one near the center would trace a greater arc length.

55. Consider two pits on a CD, one close to the center and one close to the outer edge. For a given angular velocity of the CD, which pit has a higher angular velocity? Which has a higher tangential velocity?
a. The point near the center would have the greater angular velocity and the point near the outer edge would have the higher linear velocity.
b. The point near the edge would have the greater angular velocity and the point near the center would have the higher linear velocity.
c. Both have the same angular velocity and the point near the outer edge would have the higher linear velocity.
d. Both have the same angular velocity and the point near the center would have the higher linear velocity.

56. What happens to tangential velocity as the radius of an object increases provided the angular velocity remains the same?
a. It increases because tangential velocity is directly proportional to the radius.
b. It increases because tangential velocity is inversely proportional to the radius.
c. It decreases because tangential velocity is directly proportional to the radius.
d. It decreases because tangential velocity is inversely proportional to the radius.

6.2 Uniform Circular Motion

57. Is an object in uniform circular motion accelerating? Why or why not?
a. Yes, because the velocity is not constant.
b. No, because the velocity is not constant.
c. Yes, because the velocity is constant.
d. No, because the velocity is constant.

58. An object is in uniform circular motion. Suppose the centripetal force was removed. In which direction would the object now travel?
a. In the direction of the centripetal force
b. In the direction opposite to the direction of the centripetal force
c. In the direction of the tangential velocity
d. In the direction opposite to the direction of the tangential velocity

59. An object undergoes uniform circular motion. If the radius of curvature and mass of the object are constant, what is the centripetal force proportional to?
a. $F_c \propto \frac{1}{v}$
b. $F_c \propto \frac{1}{v^2}$
c. $F_c \propto v$
d. $F_c \propto v^2$

6.3 Rotational Motion

60. Why do tornadoes produce more wind speed at the

bottom of the funnel?

a. Wind speed is greater at the bottom because rate of rotation increases as the radius increases.
b. Wind speed is greater at the bottom because rate of rotation increases as the radius decreases.
c. Wind speed is greater at the bottom because rate of rotation decreases as the radius increases.
d. Wind speed is greater at the bottom because rate of rotation decreases as the radius increases.

61. How can you maximize the torque applied to a given lever arm without applying more force?

a. The force should be applied perpendicularly to the lever arm as close as possible from the pivot point.
b. The force should be applied perpendicularly to the lever arm as far as possible from the pivot point.
c. The force should be applied parallel to the lever arm as far as possible from the pivot point.
d. The force should be applied parallel to the lever arm as close as possible from the pivot point.

62. When will an object continue spinning at the same angular velocity?

a. When net torque acting on it is zero
b. When net torque acting on it is non zero
c. When angular acceleration is positive
d. When angular acceleration is negative

CHAPTER 7
Newton's Law of Gravitation

Figure 7.1 Johannes Kepler (left) showed how the planets move, and Isaac Newton (right) discovered that gravitational force caused them to move that way. ((left) unknown, Public Domain; (right) Sir Godfrey Kneller, Public Domain)

Chapter Outline

7.1 Kepler's Laws of Planetary Motion

7.2 Newton's Law of Universal Gravitation and Einstein's Theory of General Relativity

INTRODUCTION What do a falling apple and the orbit of the moon have in common? You will learn in this chapter that each is caused by gravitational force. The motion of all celestial objects, in fact, is determined by the gravitational force, which depends on their mass and separation.

Johannes Kepler discovered three laws of planetary motion that all orbiting planets and moons follow. Years later, Isaac Newton found these laws useful in developing his law of universal gravitation. This law relates gravitational force to the masses of objects and the distance between them. Many years later still, Albert Einstein showed there was a little more to the gravitation story when he published his theory of general relativity.

7.1 Kepler's Laws of Planetary Motion

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Explain Kepler's three laws of planetary motion
- Apply Kepler's laws to calculate characteristics of orbits

Section Key Terms

aphelion	Copernican model	eccentricity
Kepler's laws of planetary motion	perihelion	Ptolemaic model

Concepts Related to Kepler's Laws of Planetary Motion

Examples of orbits abound. Hundreds of artificial satellites orbit Earth together with thousands of pieces of debris. The moon's orbit around Earth has intrigued humans from time immemorial. The orbits of planets, asteroids, meteors, and comets around the sun are no less interesting. If we look farther, we see almost unimaginable numbers of stars, galaxies, and other celestial objects orbiting one another and interacting through gravity.

All these motions are governed by gravitational force. The orbital motions of objects in our own solar system are simple enough to describe with a few fairly simple laws. The orbits of planets and moons satisfy the following two conditions:

- The mass of the orbiting object, m, is small compared to the mass of the object it orbits, M.
- The system is isolated from other massive objects.

Based on the motion of the planets about the sun, Kepler devised a set of three classical laws, called **Kepler's laws of planetary motion**, that describe the orbits of all bodies satisfying these two conditions:

1. The orbit of each planet around the sun is an ellipse with the sun at one focus.
2. Each planet moves so that an imaginary line drawn from the sun to the planet sweeps out equal areas in equal times.
3. The ratio of the squares of the periods of any two planets about the sun is equal to the ratio of the cubes of their average distances from the sun.

These descriptive laws are named for the German astronomer Johannes Kepler (1571–1630). He devised them after careful study (over some 20 years) of a large amount of meticulously recorded observations of planetary motion done by Tycho Brahe (1546–1601). Such careful collection and detailed recording of methods and data are hallmarks of good science. Data constitute the evidence from which new interpretations and meanings can be constructed. Let's look closer at each of these laws.

Kepler's First Law

The orbit of each planet about the sun is an ellipse with the sun at one focus, as shown in Figure 7.2. The planet's closest approach to the sun is called **aphelion** and its farthest distance from the sun is called **perihelion**.

Figure 7.2 (a) An ellipse is a closed curve such that the sum of the distances from a point on the curve to the two foci (f_1 and f_2) is constant. (b) For any closed orbit, *m* follows an elliptical path with *M* at one focus. (c) The aphelion (*r*a) is the closest distance between the planet and the sun, while the perihelion (*r*p) is the farthest distance from the sun.

If you know the aphelion (r_a) and perihelion (r_p) distances, then you can calculate the semi-major axis (*a*) and semi-minor axis (*b*).

$$a = \frac{(r_a + r_p)}{2}$$
$$b = \sqrt{r_a r_p}$$

Figure 7.3 You can draw an ellipse as shown by putting a pin at each focus, and then placing a loop of string around a pen and the pins and tracing a line on the paper.

Kepler's Second Law

Each planet moves so that an imaginary line drawn from the sun to the planet sweeps out equal areas in equal times, as shown in Figure 7.4.

Figure 7.4 The shaded regions have equal areas. The time for *m* to go from A to B is the same as the time to go from C to D and from E to F. The mass *m* moves fastest when it is closest to *M*. Kepler's second law was originally devised for planets orbiting the sun, but it has broader validity.

TIPS FOR SUCCESS

Note that while, for historical reasons, Kepler's laws are stated for planets orbiting the sun, they are actually valid for all bodies satisfying the two previously stated conditions.

Kepler's Third Law

The ratio of the periods squared of any two planets around the sun is equal to the ratio of their average distances from the sun cubed. In equation form, this is

$$\frac{T_1^2}{T_2^2} = \frac{r_1^3}{r_2^3},$$

where T is the period (time for one orbit) and r is the average distance (also called orbital radius). This equation is valid only for comparing two small masses orbiting a single large mass. Most importantly, this is only a descriptive equation; it gives no information about the cause of the equality.

LINKS TO PHYSICS

History: Ptolemy vs. Copernicus

Before the discoveries of Kepler, Copernicus, Galileo, Newton, and others, the solar system was thought to revolve around Earth as shown in Figure 7.5 **(a)**. This is called the **Ptolemaic model**, named for the Greek philosopher Ptolemy who lived in the second century AD. The Ptolemaic model is characterized by a list of facts for the motions of planets, with no explanation of cause and effect. There tended to be a different rule for each heavenly body and a general lack of simplicity.

Figure 7.5 **(b)** represents the modern or **Copernican model**. In this model, a small set of rules and a single underlying force explain not only all planetary motion in the solar system, but also all other situations involving gravity. The breadth and simplicity of the laws of physics are compelling.

Figure 7.5 (a) The Ptolemaic model of the universe has Earth at the center with the moon, the planets, the sun, and the stars revolving about it in complex circular paths. This geocentric (Earth-centered) model, which can be made progressively more accurate by adding more circles, is purely descriptive, containing no hints about the causes of these motions. (b) The Copernican heliocentric (sun-centered) model is a simpler and more accurate model.

Nicolaus Copernicus (1473–1543) first had the idea that the planets circle the sun, in about 1514. It took him almost 20 years to work out the mathematical details for his model. He waited another 10 years or so to publish his work. It is thought he hesitated because he was afraid people would make fun of his theory. Actually, the reaction of many people was more one of fear and anger. Many people felt the Copernican model threatened their basic belief system. About 100 years later, the astronomer Galileo was put under house arrest for providing evidence that planets, including Earth, orbited the sun. In all, it took almost 300 years for everyone to admit that Copernicus had been right all along.

GRASP CHECK

Explain why Earth does actually appear to be the center of the solar system.

a. Earth appears to be the center of the solar system because Earth is at the center of the universe, and everything revolves around it in a circular orbit.
b. Earth appears to be the center of the solar system because, in the reference frame of Earth, the sun, moon, and planets all appear to move across the sky as if they were circling Earth.
c. Earth appears to be at the center of the solar system because Earth is at the center of the solar system and all the heavenly bodies revolve around it.
d. Earth appears to be at the center of the solar system because Earth is located at one of the foci of the elliptical orbit of the sun, moon, and other planets.

Virtual Physics

Acceleration

This simulation allows you to create your own solar system so that you can see how changing distances and masses determines the orbits of planets. Click *Help* for instructions.

Click to view content (https://archive.cnx.org/specials/ee816dff-0b5f-4e6f-8250-f9fb9e39d716/my-solar-system/)

GRASP CHECK

When the central object is off center, how does the speed of the orbiting object vary?

a. The orbiting object moves fastest when it is closest to the central object and slowest when it is farthest away.
b. The orbiting object moves slowest when it is closest to the central object and fastest when it is farthest away.

c. The orbiting object moves with the same speed at every point on the circumference of the elliptical orbit.
d. There is no relationship between the speed of the object and the location of the planet on the circumference of the orbit.

Calculations Related to Kepler's Laws of Planetary Motion

Kepler's First Law

Refer back to Figure 7.2 **(a)**. Notice which distances are constant. The foci are fixed, so distance $\overline{f_1 f_2}$ is a constant. The definition of an ellipse states that the sum of the distances $\overline{f_1 m} + \overline{m f_2}$ is also constant. These two facts taken together mean that the perimeter of triangle $\Delta f_1 m f_2$ must also be constant. Knowledge of these constants will help you determine positions and distances of objects in a system that includes one object orbiting another.

Kepler's Second Law

Refer back to Figure 7.4. The second law says that the segments have equal area and that it takes equal time to sweep through each segment. That is, the time it takes to travel from A to B equals the time it takes to travel from C to D, and so forth. Velocity **v** equals distance *d* divided by time *t*: $\mathbf{v} = d/t$. Then, $t = d/\mathbf{v}$, so distance divided by velocity is also a constant. For example, if we know the average velocity of Earth on June 21 and December 21, we can compare the distance Earth travels on those days.

The degree of elongation of an elliptical orbit is called its **eccentricity** (*e*). Eccentricity is calculated by dividing the distance *f* from the center of an ellipse to one of the foci by half the long axis *a*.

$$(e) = f/a \tag{7.1}$$

When $e = 0$, the ellipse is a circle.

The area of an ellipse is given by $A = \pi ab$, where *b* is half the short axis. If you know the axes of Earth's orbit and the area Earth sweeps out in a given period of time, you can calculate the fraction of the year that has elapsed.

WORKED EXAMPLE

Kepler's First Law

At its closest approach, a moon comes within 200,000 km of the planet it orbits. At that point, the moon is 300,000 km from the other focus of its orbit, f_2. The planet is focus f_1 of the moon's elliptical orbit. How far is the moon from the planet when it is 260,000 km from f_2?

Strategy

Show and label the ellipse that is the orbit in your solution. Picture the triangle $f_1 m f_2$ collapsed along the major axis and add up the lengths of the three sides. Find the length of the unknown side of the triangle when the moon is 260,000 km from f_2.

Solution

Perimeter of $f_1 m f_2 = 200,000 \text{ km} + 100,000 \text{ km} + 300,000 \text{ km} = 600,000 \text{ km}.$

$$mf_1 = 600,000 \text{ km} - (100,000 \text{ km} + 200,000 \text{ km}) = 240,000 \text{ km}.$$

Discussion

The perimeter of triangle $f_1 m f_2$ must be constant because the distance between the foci does not change and Kepler's first law says the orbit is an ellipse. For any ellipse, the sum of the two sides of the triangle, which are $f_1 m$ and mf_2, is constant.

WORKED EXAMPLE

Kepler's Second Law

Figure 7.6 shows the major and minor axes of an ellipse. The semi-major and semi-minor axes are half of these, respectively.

Figure 7.6 The major axis is the length of the ellipse, and the minor axis is the width of the ellipse. The semi-major axis is half the major axis, and the semi-minor axis is half the minor axis.

Earth's orbit is slightly elliptical, with a semi-major axis of 152 million km and a semi-minor axis of 147 million km. If Earth's period is 365.26 days, what area does an Earth-to-sun line sweep past in one day?

Strategy

Each day, Earth sweeps past an equal-sized area, so we divide the total area by the number of days in a year to find the area swept past in one day. For total area use $A = \pi ab$. Calculate A, the area inside Earth's orbit and divide by the number of days in a year (i.e., its period).

Solution

$$\begin{aligned} \text{area per day} &= \frac{\text{total area}}{\text{total number of days}} \\ &= \frac{\pi ab}{365\text{ d}} \\ &= \frac{\pi(1.47\times10^{8}\text{ km})(1.52\times10^{3}\text{ km})}{365\text{ d}} \\ &= 1.92\times10^{14}\text{ km}^2/\text{d} \end{aligned} \tag{7.2}$$

The area swept out in one day is thus $1.92 \times 10^{14}\text{km}^2$.

Discussion

The answer is based on Kepler's law, which states that a line from a planet to the sun sweeps out equal areas in equal times.

Kepler's Third Law

Kepler's third law states that the ratio of the squares of the periods of any two planets (T_1, T_2) is equal to the ratio of the cubes of their average orbital distance from the sun (r_1, r_2). Mathematically, this is represented by

$$\frac{T_1^2}{T_2^2} = \frac{r_1^3}{r_2^3}.$$

From this equation, it follows that the ratio r^3/T^2 is the same for all planets in the solar system. Later we will see how the work of Newton leads to a value for this constant.

WORKED EXAMPLE

Kepler's Third Law

Given that the moon orbits Earth each 27.3 days and that it is an average distance of $3.84 \times 10^8\text{m}$ from the center of Earth, calculate the period of an artificial satellite orbiting at an average altitude of 1,500 km above Earth's surface.

Strategy

The period, or time for one orbit, is related to the radius of the orbit by Kepler's third law, given in mathematical form by

$\frac{T_1^2}{T_2^2} = \frac{r_1^3}{r_2^3}$. Let us use the subscript 1 for the moon and the subscript 2 for the satellite. We are asked to find T_2. The given information tells us that the orbital radius of the moon is $r_1 = 3.84 \times 10^8 \text{ m}$, and that the period of the moon is $T_1 = 27.3 \text{ days}$. The height of the artificial satellite above Earth's surface is given, so to get the distance r_2 from the center of Earth we must add the height to the radius of Earth (6380 km). This gives $r_2 = 1500 \text{ km} + 6380 \text{ km} = 7880 \text{ km}$. Now all quantities are known, so T_2 can be found.

Solution

To solve for T_2, we cross-multiply and take the square root, yielding

$$T_2^2 = T_1^2 \left(\frac{r_2}{r_1}\right)^3; \; T_2 = T_1 \left(\frac{r_2}{r_1}\right)^{\frac{3}{2}}$$

$$T_2 = (27.3 \text{ d}) \left(\frac{24.0 \text{ h}}{\text{d}}\right) \left(\frac{7880 \text{ km}}{3.84 \times 10^5 \text{ km}}\right)^{\frac{3}{2}} = 1.93 \text{ h}. \tag{7.3}$$

Discussion

This is a reasonable period for a satellite in a fairly low orbit. It is interesting that any satellite at this altitude will complete one orbit in the same amount of time.

Practice Problems

1. A planet with no axial tilt is located in another solar system. It circles its sun in a very elliptical orbit so that the temperature varies greatly throughout the year. If the year there has 612 days and the inhabitants celebrate the coldest day on day 1 of their calendar, when is the warmest day?
   a. Day 1
   b. Day 153
   c. Day 306
   d. Day 459

2. A geosynchronous Earth satellite is one that has an orbital period of precisely 1 day. Such orbits are useful for communication and weather observation because the satellite remains above the same point on Earth (provided it orbits in the equatorial plane in the same direction as Earth's rotation). The ratio $\frac{r^3}{T^2}$ for the moon is $1.01 \times 10^{18} \frac{\text{km}^3}{y^2}$. Calculate the radius of the orbit of such a satellite.
   a. $2.75 \times 10^3 \text{ km}$
   b. $1.96 \times 10^4 \text{ km}$
   c. $1.40 \times 10^5 \text{ km}$
   d. $1.00 \times 10^6 \text{ km}$

Check Your Understanding

3. Are Kepler's laws purely descriptive, or do they contain causal information?
   a. Kepler's laws are purely descriptive.
   b. Kepler's laws are purely causal.
   c. Kepler's laws are descriptive as well as causal.
   d. Kepler's laws are neither descriptive nor causal.

4. True or false—According to Kepler's laws of planetary motion, a satellite increases its speed as it approaches its parent body and decreases its speed as it moves away from the parent body.
   a. True
   b. False

5. Identify the locations of the foci of an elliptical orbit.
   a. One focus is the parent body, and the other is located at the opposite end of the ellipse, at the same distance from the center as the parent body.
   b. One focus is the parent body, and the other is located at the opposite end of the ellipse, at half the distance from the center as the parent body.

c. One focus is the parent body and the other is located outside of the elliptical orbit, on the line on which is the semi-major axis of the ellipse.
d. One focus is on the line containing the semi-major axis of the ellipse, and the other is located anywhere on the elliptical orbit of the satellite.

7.2 Newton's Law of Universal Gravitation and Einstein's Theory of General Relativity

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Explain Newton's law of universal gravitation and compare it to Einstein's theory of general relativity
- Perform calculations using Newton's law of universal gravitation

Section Key Terms

Einstein's theory of general relativity	gravitational constant	Newton's universal law of gravitation

Concepts Related to Newton's Law of Universal Gravitation

Sir Isaac Newton was the first scientist to precisely define the gravitational force, and to show that it could explain both falling bodies and astronomical motions. See Figure 7.7. But Newton was not the first to suspect that the same force caused both our weight and the motion of planets. His forerunner, Galileo Galilei, had contended that falling bodies and planetary motions had the same cause. Some of Newton's contemporaries, such as Robert Hooke, Christopher Wren, and Edmund Halley, had also made some progress toward understanding gravitation. But Newton was the first to propose an exact mathematical form and to use that form to show that the motion of heavenly bodies should be conic sections—circles, ellipses, parabolas, and hyperbolas. This theoretical prediction was a major triumph. It had been known for some time that moons, planets, and comets follow such paths, but no one had been able to propose an explanation of the mechanism that caused them to follow these paths and not others.

Figure 7.7 The popular legend that Newton suddenly discovered the law of universal gravitation when an apple fell from a tree and hit him on the head has an element of truth in it. A more probable account is that he was walking through an orchard and wondered why all the apples fell in the same direction with the same acceleration. Great importance is attached to it because Newton's universal law of gravitation and his laws of motion answered very old questions about nature and gave tremendous support to the notion of underlying simplicity and unity in nature. Scientists still expect underlying simplicity to emerge from their ongoing inquiries into nature.

The gravitational force is relatively simple. It is always attractive, and it depends only on the masses involved and the distance

between them. Expressed in modern language, **Newton's universal law of gravitation** states that every object in the universe attracts every other object with a force that is directed along a line joining them. The force is directly proportional to the product of their masses and inversely proportional to the square of the distance between them. This attraction is illustrated by Figure 7.8.

Figure 7.8 Gravitational attraction is along a line joining the centers of mass (CM) of the two bodies. The magnitude of the force on each body is the same, consistent with Newton's third law (action-reaction).

For two bodies having masses m and M with a distance r between their centers of mass, the equation for Newton's universal law of gravitation is

$$\mathbf{F} = G\frac{mM}{r^2}$$

where **F** is the magnitude of the gravitational force and G is a proportionality factor called the **gravitational constant**. G is a universal constant, meaning that it is thought to be the same everywhere in the universe. It has been measured experimentally to be $G = 6.673 \times 10^{-11}\,\mathrm{N \cdot m^2/kg^2}$.

If a person has a mass of 60.0 kg, what would be the force of gravitational attraction on him at Earth's surface? G is given above, Earth's mass M is 5.97×10^{24} kg, and the radius r of Earth is 6.38×10^6 m. Putting these values into Newton's universal law of gravitation gives

$$\mathbf{F} = G\frac{mM}{r^2} = \left(6.673 \times 10^{-11}\,\frac{\mathrm{N \cdot m^2}}{\mathrm{kg^2}}\right)\left(\frac{(60.0\ \mathrm{kg})\left(5.97 \times 10^{24}\ \mathrm{kg}\right)}{\left(6.38 \times 10^6\ \mathrm{m}\right)^2}\right) = 584\ \ \mathrm{N}$$

We can check this result with the relationship: $\mathbf{F} = m\mathbf{g} = (60\ \mathrm{kg})(9.8\mathrm{m/s^2}) = 588\ \mathrm{N}$

You may remember that **g**, the acceleration due to gravity, is another important constant related to gravity. By substituting **g** for **a** in the equation for Newton's second law of motion we get $\mathbf{F} = m\mathbf{g}$. Combining this with the equation for universal gravitation gives

$$m\mathbf{g} = G\frac{mM}{r^2}$$

Cancelling the mass m on both sides of the equation and filling in the values for the gravitational constant and mass and radius of the Earth, gives the value of g, which may look familiar.

$$\mathbf{g} = G\frac{M}{r^2} = \left(6.67 \times 10^{-11}\,\frac{\mathrm{N \cdot m^2}}{\mathrm{kg^2}}\right)\left(\frac{5.98 \times 10^{24}\mathrm{kg}}{\left(6.38 \times 10^6\mathrm{m}\right)^2}\right) = 9.80\mathrm{m/s^2}$$

This is a good point to recall the difference between mass and weight. Mass is the amount of matter in an object; weight is the

force of attraction between the mass within two objects. Weight can change because g is different on every moon and planet. An object's mass m does not change but its weight $m\mathbf{g}$ can.

Virtual Physics

Gravity and Orbits

Move the sun, Earth, moon and space station in this simulation to see how it affects their gravitational forces and orbital paths. Visualize the sizes and distances between different heavenly bodies. Turn off gravity to see what would happen without it!

Click to view content (https://archive.cnx.org/specials/a14085c8-96b8-4d04-bb5a-56d9ccbe6e69/gravity-and-orbits/)

GRASP CHECK

Why doesn't the Moon travel in a smooth circle around the Sun?

a. The Moon is not affected by the gravitational field of the Sun.
b. The Moon is not affected by the gravitational field of the Earth.
c. The Moon is affected by the gravitational fields of both the Earth and the Sun, which are always additive.
d. The moon is affected by the gravitational fields of both the Earth and the Sun, which are sometimes additive and sometimes opposite.

Snap Lab

Take-Home Experiment: Falling Objects

In this activity you will study the effects of mass and air resistance on the acceleration of falling objects. Make predictions (hypotheses) about the outcome of this experiment. Write them down to compare later with results.

- Four sheets of $8\text{-}1/2 \times 11$-inch paper

Procedure

- Take four identical pieces of paper.
  - Crumple one up into a small ball.
  - Leave one uncrumpled.
  - Take the other two and crumple them up together, so that they make a ball of exactly twice the mass of the other crumpled ball.
  - Now compare which ball of paper lands first when dropped simultaneously from the same height.
    1. Compare crumpled one-paper ball with crumpled two-paper ball.
    2. Compare crumpled one-paper ball with uncrumpled paper.

GRASP CHECK

Why do some objects fall faster than others near the surface of the earth if all mass is attracted equally by the force of gravity?

a. Some objects fall faster because of air resistance, which acts in the direction of the motion of the object and exerts more force on objects with less surface area.
b. Some objects fall faster because of air resistance, which acts in the direction opposite the motion of the object and exerts more force on objects with less surface area.
c. Some objects fall faster because of air resistance, which acts in the direction of motion of the object and exerts more force on objects with more surface area.
d. Some objects fall faster because of air resistance, which acts in the direction opposite the motion of the object and exerts more force on objects with more surface area.

It is possible to derive Kepler's third law from Newton's law of universal gravitation. Applying Newton's second law of motion to

angular motion gives an expression for centripetal force, which can be equated to the expression for force in the universal gravitation equation. This expression can be manipulated to produce the equation for Kepler's third law. We saw earlier that the expression r^3/T^2 is a constant for satellites orbiting the same massive object. The derivation of Kepler's third law from Newton's law of universal gravitation and Newton's second law of motion yields that constant:

$$\frac{r^3}{T^2} = \frac{GM}{4\pi^2}$$

where M is the mass of the central body about which the satellites orbit (for example, the sun in our solar system). The usefulness of this equation will be seen later.

The universal gravitational constant G is determined experimentally. This definition was first done accurately in 1798 by English scientist Henry Cavendish (1731–1810), more than 100 years after Newton published his universal law of gravitation. The measurement of G is very basic and important because it determines the strength of one of the four forces in nature. Cavendish's experiment was very difficult because he measured the tiny gravitational attraction between two ordinary-sized masses (tens of kilograms at most) by using an apparatus like that in Figure 7.9. Remarkably, his value for G differs by less than 1% from the modern value.

Figure 7.9 Cavendish used an apparatus like this to measure the gravitational attraction between two suspended spheres (m) and two spheres on a stand (M) by observing the amount of torsion (twisting) created in the fiber. The distance between the masses can be varied to check the dependence of the force on distance. Modern experiments of this type continue to explore gravity.

Einstein's Theory of General Relativity

Einstein's theory of general relativity explained some interesting properties of gravity not covered by Newton's theory. Einstein based his theory on the postulate that acceleration and gravity have the same effect and cannot be distinguished from each other. He concluded that light must fall in both a gravitational field and in an accelerating reference frame. Figure 7.10 shows this effect (greatly exaggerated) in an accelerating elevator. In Figure 7.10**(a)**, the elevator accelerates upward in zero gravity. In Figure 7.10**(b)**, the room is not accelerating but is subject to gravity. The effect on light is the same: it "falls" downward in both situations. The person in the elevator cannot tell whether the elevator is accelerating in zero gravity or is stationary and subject to gravity. Thus, gravity affects the path of light, even though we think of gravity as acting between masses, while photons are massless.

Figure 7.10 (a) A beam of light emerges from a flashlight in an upward-accelerating elevator. Since the elevator moves up during the time the light takes to reach the wall, the beam strikes lower than it would if the elevator were not accelerated. (b) Gravity must have the same effect on light, since it is not possible to tell whether the elevator is accelerating upward or is stationary and acted upon by gravity.

Einstein's theory of general relativity got its first verification in 1919 when starlight passing near the sun was observed during a solar eclipse. (See Figure 7.11.) During an eclipse, the sky is darkened and we can briefly see stars. Those on a line of sight nearest the sun should have a shift in their apparent positions. Not only was this shift observed, but it agreed with Einstein's predictions well within experimental uncertainties. This discovery created a scientific and public sensation. Einstein was now a folk hero as well as a very great scientist. The bending of light by matter is equivalent to a bending of space itself, with light following the curve. This is another radical change in our concept of space and time. It is also another connection that any particle with mass *or* energy (e.g., massless photons) is affected by gravity.

Figure 7.11 This schematic shows how light passing near a massive body like the sun is curved toward it. The light that reaches the Earth then seems to be coming from different locations than the known positions of the originating stars. Not only was this effect observed, but the amount of bending was precisely what Einstein predicted in his general theory of relativity.

To summarize the two views of gravity, Newton envisioned gravity as a tug of war along the line connecting any two objects in the universe. In contrast, Einstein envisioned gravity as a bending of space-time by mass.

BOUNDLESS PHYSICS

NASA gravity probe B

NASA's Gravity Probe B (GP-B) mission has confirmed two key predictions derived from Albert Einstein's general theory of relativity. The probe, shown in Figure 7.12 was launched in 2004. It carried four ultra-precise gyroscopes designed to measure two effects hypothesized by Einstein's theory:

- The geodetic effect, which is the warping of space and time by the gravitational field of a massive body (in this case, Earth)
- The frame-dragging effect, which is the amount by which a spinning object pulls space and time with it as it rotates

Figure 7.12 Artist concept of Gravity Probe B spacecraft in orbit around the Earth. (credit: NASA/MSFC)

Both effects were measured with unprecedented precision. This was done by pointing the gyroscopes at a single star while orbiting Earth in a polar orbit. As predicted by relativity theory, the gyroscopes experienced very small, but measureable, changes in the direction of their spin caused by the pull of Earth's gravity.

The principle investigator suggested imagining Earth spinning in honey. As Earth rotates it drags space and time with it as it would a surrounding sea of honey.

GRASP CHECK

According to the general theory of relativity, a gravitational field bends light. What does this have to do with time and space?

a. Gravity has no effect on the space-time continuum, and gravity only affects the motion of light.
b. The space-time continuum is distorted by gravity, and gravity has no effect on the motion of light.
c. Gravity has no effect on either the space-time continuum or on the motion of light.
d. The space-time continuum is distorted by gravity, and gravity affects the motion of light.

Calculations Based on Newton's Law of Universal Gravitation

TIPS FOR SUCCESS

When performing calculations using the equations in this chapter, use units of kilograms for mass, meters for distances, newtons for force, and seconds for time.

The mass of an object is constant, but its weight varies with the strength of the gravitational field. This means the value of $\mathbf{g}$ varies from place to place in the universe. The relationship between force, mass, and acceleration from the second law of motion can be written in terms of $\mathbf{g}$.

$$\mathbf{F} = m\mathbf{a} = m\mathbf{g}$$

In this case, the force is the weight of the object, which is caused by the gravitational attraction of the planet or moon on which the object is located. We can use this expression to compare weights of an object on different moons and planets.

WATCH PHYSICS

Mass and Weight Clarification

This video shows the mathematical basis of the relationship between mass and weight. The distinction between mass and weight are clearly explained. The mathematical relationship between mass and weight are shown mathematically in terms of the equation for Newton's law of universal gravitation and in terms of his second law of motion.

Click to view content (https://www.khanacademy.org/embed_video?v=IuBoeDihLUc)

GRASP CHECK

Would you have the same mass on the moon as you do on Earth? Would you have the same weight?

a. You would weigh more on the moon than on Earth because gravity on the moon is stronger than gravity on Earth.
b. You would weigh less on the moon than on Earth because gravity on the moon is weaker than gravity on Earth.
c. You would weigh less on the moon than on Earth because gravity on the moon is stronger than gravity on Earth.
d. You would weigh more on the moon than on Earth because gravity on the moon is weaker than gravity on Earth.

Two equations involving the gravitational constant, G, are often useful. The first is Newton's equation, $\mathbf{F} = G\frac{mM}{r^2}$. Several of the values in this equation are either constants or easily obtainable. $\mathbf{F}$ is often the weight of an object on the surface of a large object with mass M, which is usually known. The mass of the smaller object, m, is often known, and G is a universal constant with the same value anywhere in the universe. This equation can be used to solve problems involving an object on or orbiting Earth or other massive celestial object. Sometimes it is helpful to equate the right-hand side of the equation to $m\mathbf{g}$ and cancel the m on both sides.

The equation $\frac{r^3}{T^2} = \frac{GM}{4\pi^2}$ is also useful for problems involving objects in orbit. Note that there is no need to know the mass of the object. Often, we know the radius r or the period T and want to find the other. If these are both known, we can use the equation to calculate the mass of a planet or star.

WATCH PHYSICS

Mass and Weight Clarification

This video demonstrates calculations involving Newton's universal law of gravitation.

Click to view content (https://www.khanacademy.org/embed_video?v=391txUI76gM)

GRASP CHECK

Identify the constants g and G.

a. g and G are both the acceleration due to gravity
b. g is acceleration due to gravity on Earth and G is the universal gravitational constant.
c. g is the gravitational constant and G is the acceleration due to gravity on Earth.
d. g and G are both the universal gravitational constant.

WORKED EXAMPLE

Change in g

The value of $\mathbf{g}$ on the planet Mars is 3.71 m/s^2. If you have a mass of 60.0 kg on Earth, what would be your mass on Mars? What would be your weight on Mars?

Strategy

Weight equals acceleration due to gravity times mass: $\mathbf{W} = m\mathbf{g}$. An object's mass is constant. Call acceleration due to gravity on Mars $\mathbf{g}_M$ and weight on Mars $\mathbf{W}_M$.

Solution

Mass on Mars would be the same, 60 kg.

$$\mathbf{W}_M = m\mathbf{g}_M = (60.0\text{ kg})\left(3.71\text{m/s}^2\right) = 223\text{ N} \tag{7.4}$$

Discussion

The value of **g** on any planet depends on the mass of the planet and the distance from its center. If the material below the surface varies from point to point, the value of **g** will also vary slightly.

WORKED EXAMPLE

Earth's g at the Moon

Find the acceleration due to Earth's gravity at the distance of the moon.

$$\begin{aligned} &G = 6.67 \times 10^{-11}\text{ N}\cdot\text{m}^2/\text{kg}^2 \\ &\text{Earth-moon distance} = 3.84 \times 10^8\text{ m} \\ &\text{Earth's mass} = 5.98 \times 10^{24}\text{ kg} \end{aligned} \tag{7.5}$$

Express the force of gravity in terms of *g*.

$$\text{F} = \text{W} = m\text{a} = m\text{g} \tag{7.6}$$

Combine with the equation for universal gravitation.

$$m\mathbf{g} = mG\frac{M}{r^2} \tag{7.7}$$

Solution

Cancel *m* and substitute.

$$\mathbf{g} = G\frac{M}{r^2} = \left(6.67\times10^{-11}\frac{\text{N}\cdot\text{m}^2}{\text{kg}^2}\right)\left(\frac{5.98\times10^{24}\text{ kg}}{\left(3.84\times10^8\text{ m}\right)^2}\right) = 2.70\times10^{-3}\text{m/s}^2 \tag{7.8}$$

Discussion

The value of **g** for the moon is 1.62 m/s². Comparing this value to the answer, we see that Earth's gravitational influence on an object on the moon's surface would be insignificant.

Practice Problems

6. What is the mass of a person who weighs 600 N?
   a. 6.00 kg
   b. 61.2 kg
   c. 600 kg
   d. 610 kg

7. Calculate Earth's mass given that the acceleration due to gravity at the North Pole is 9.830m/s^2 and the radius of the Earth is 6371 km from pole to center.
   a. 5.94×10^{17} kg
   b. 5.94×10^{24} kg
   c. 9.36×10^{17} kg
   d. 9.36×10^{24} kg

Check Your Understanding

8. Some of Newton's predecessors and contemporaries also studied gravity and proposed theories. What important advance did Newton make in the study of gravity that the other scientists had failed to do?
   a. He gave an exact mathematical form for the theory.

b. He added a correction term to a previously existing formula.
c. Newton found the value of the universal gravitational constant.
d. Newton showed that gravitational force is always attractive.

9. State the law of universal gravitation in words only.
a. Gravitational force between two objects is directly proportional to the sum of the squares of their masses and inversely proportional to the square of the distance between them.
b. Gravitational force between two objects is directly proportional to the product of their masses and inversely proportional to the square of the distance between them.
c. Gravitational force between two objects is directly proportional to the sum of the squares of their masses and inversely proportional to the distance between them.
d. Gravitational force between two objects is directly proportional to the product of their masses and inversely proportional to the distance between them.

10. Newton's law of universal gravitation explains the paths of what?
a. A charged particle
b. A ball rolling on a plane surface
c. A planet moving around the sun
d. A stone tied to a string and whirled at constant speed in a horizontal circle

KEY TERMS

aphelion closest distance between a planet and the sun (called apoapsis for other celestial bodies)

Copernican model the model of the solar system where the sun is at the center of the solar system and all the planets orbit around it; this is also called the heliocentric model

eccentricity a measure of the separation of the foci of an ellipse

Einstein's theory of general relativity the theory that gravitational force results from the bending of spacetime by an object's mass

gravitational constant the proportionality constant in Newton's law of universal gravitation

Kepler's laws of planetary motion three laws derived by Johannes Kepler that describe the properties of all orbiting satellites

Newton's universal law of gravitation states that gravitational force between two objects is directly proportional to the product of their masses and inversely proportional to the square of the distance between them.

perihelion farthest distance between a planet and the sun (called periapsis for other celestial bodies)

Ptolemaic model the model of the solar system where Earth is at the center of the solar system and the sun and all the planets orbit around it; this is also called the geocentric model

SECTION SUMMARY

7.1 Kepler's Laws of Planetary Motion

- All satellites follow elliptical orbits.
- The line from the satellite to the parent body sweeps out equal areas in equal time.
- The radius cubed divided by the period squared is a constant for all satellites orbiting the same parent body.

7.2 Newton's Law of Universal Gravitation and Einstein's Theory of General Relativity

- Newton's law of universal gravitation provides a mathematical basis for gravitational force and Kepler's laws of planetary motion.
- Einstein's theory of general relativity shows that gravitational fields change the path of light and warp space and time.
- An object's mass is constant, but its weight changes when acceleration due to gravity, **g**, changes.

KEY EQUATIONS

7.1 Kepler's Laws of Planetary Motion

Kepler's third law	$\frac{T_1^2}{T_2^2} = \frac{r_1^3}{r_2^3}$
eccentricity	$e = \frac{f}{a}$
area of an ellipse	$A = \pi ab$
semi-major axis of an ellipse	$a = \left(r_\text{a} + r_\text{p}\right)/2$
semi-minor axis of an ellipse	$b = \sqrt{r_\text{a} r_\text{p}}$

7.2 Newton's Law of Universal Gravitation and Einstein's Theory of General Relativity

Newton's second law of motion	$\mathbf{F} = m\mathbf{a} = m\mathbf{g}$
Newton's universal law of gravitation	$\mathbf{F} = G\frac{mM}{r^2}$
acceleration due to gravity	$\mathbf{g} = G\frac{M}{r^2}$
constant for satellites orbiting the same massive object	$\frac{r^3}{T^2} = \frac{GM}{4\pi^2}$

CHAPTER REVIEW

Concept Items

7.1 Kepler's Laws of Planetary Motion

1. A circle is a special case of an ellipse. Explain how a circle is different from other ellipses.
   a. The foci of a circle are at the same point and are located at the center of the circle.
   b. The foci of a circle are at the same point and are

located at the circumference of the circle.
   c. The foci of a circle are at the same point and are located outside of the circle.
   d. The foci of a circle are at the same point and are located anywhere on the diameter, except on its midpoint.

2. Comets have very elongated elliptical orbits with the sun at one focus. Using Kepler's Law, explain why a comet travels much faster near the sun than it does at the other end of the orbit.
   a. Because the satellite sweeps out equal areas in equal times
   b. Because the satellite sweeps out unequal areas in equal times
   c. Because the satellite is at the other focus of the ellipse
   d. Because the square of the period of the satellite is proportional to the cube of its average distance from the sun

3. True or False—A planet-satellite system must be isolated from other massive objects to follow Kepler's laws of planetary motion.
   a. True
   b. False

4. Explain why the string, pins, and pencil method works for drawing an ellipse.
   a. The string, pins, and pencil method works because the length of the two sides of the triangle remains constant as you are drawing the ellipse.
   b. The string, pins, and pencil method works because the area of the triangle remains constant as you are drawing the ellipse.
   c. The string, pins, and pencil method works because the perimeter of the triangle remains constant as you are drawing the ellipse.
   d. The string, pins, and pencil method works because the volume of the triangle remains constant as you are drawing the ellipse.

7.2 Newton's Law of Universal Gravitation and Einstein's Theory of General Relativity

5. Describe the postulate on which Einstein based the theory of general relativity and describe an everyday experience that illustrates this postulate.
   a. Gravity and velocity have the same effect and cannot be distinguished from each other. An acceptable illustration of this is any description of the feeling of constant velocity in a situation where no outside frame of reference is considered.
   b. Gravity and velocity have different effects and can be distinguished from each other. An acceptable illustration of this is any description of the feeling of constant velocity in a situation where no outside frame of reference is considered.
   c. Gravity and acceleration have the same effect and cannot be distinguished from each other. An acceptable illustration of this is any description of the feeling of acceleration in a situation where no outside frame of reference is considered.
   d. Gravity and acceleration have different effects and can be distinguished from each other. An acceptable illustration of this is any description of the feeling of acceleration in a situation where no outside frame of reference is considered.

6. Titan, with a radius of 2.58×10^6 m, is the largest moon of the planet Saturn. If the mass of Titan is 1.35×10^{23} kg, what is the acceleration due to gravity on the surface of this moon?
   a. $1.35\ \text{m/s}^2$
   b. $3.49\ \text{m/s}^2$
   c. $3.49 \times 10^6\ \text{m/s}^2$
   d. $1.35 \times 10^6\ \text{m/s}^2$

7. Saturn's moon Titan has an orbital period of 15.9 days. If Saturn has a mass of 5.68×10^{23} kg, what is the average distance from Titan to the center of Saturn?
   a. 1.22×10^6 m
   b. 4.26×10^7 m
   c. 5.25×10^4 km
   d. 4.26×10^{10} km

8. Explain why doubling the mass of an object doubles its weight, but doubling its distance from the center of Earth reduces its weight fourfold.
   a. The weight is two times the gravitational force between the object and Earth.
   b. The weight is half the gravitational force between the object and Earth.
   c. The weight is equal to the gravitational force between the object and Earth, and the gravitational force is inversely proportional to the distance squared between the object and Earth.
   d. The weight is directly proportional to the square of the gravitational force between the object and Earth.

9. Explain why a star on the other side of the Sun might appear to be in a location that is not its true location.
   a. It can be explained by using the concept of atmospheric refraction.
   b. It can be explained by using the concept of the special theory of relativity.
   c. It can be explained by using the concept of the general theory of relativity.
   d. It can be explained by using the concept of light

scattering in the atmosphere.

10. The Cavendish experiment marked a milestone in the study of gravity.
Part A. What important value did the experiment determine?
Part B. Why was this so difficult in terms of the masses used in the apparatus and the strength of the gravitational force?
   a. Part A. The experiment measured the acceleration due to gravity, g. Part B. Gravity is a very weak force but despite this limitation, Cavendish was able to measure the attraction between very massive objects.
   b. Part A. The experiment measured the gravitational constant, G. Part B. Gravity is a very weak force but, despite this limitation, Cavendish was able to measure the attraction between very massive objects.
   c. Part A. The experiment measured the acceleration due to gravity, g. Part B. Gravity is a very weak force but despite this limitation, Cavendish was able to measure the attraction between less massive objects.
   d. Part A. The experiment measured the gravitational constant, G. Part B. Gravity is a very weak force but despite this limitation, Cavendish was able to measure the attraction between less massive objects.

Critical Thinking Items

7.1 Kepler's Laws of Planetary Motion

11. In the figure, the time it takes for the planet to go from A to B, C to D, and E to F is the same.

Compare the areas A_1, A_2, and A_3 in terms of size.
   a. $A_1 \neq A_2 \neq A_3$
   b. $A_1 = A_2 = A_3$
   c. $A_1 = A_2 > A_3$
   d. $A_1 > A_2 = A_3$

12. A moon orbits a planet in an elliptical orbit. The foci of the ellipse are 50, 000 km apart. The closest approach of the moon to the planet is 400, 000 km. What is the length of the major axis of the orbit?
   a. 400, 000 km
   b. 450, 000, km
   c. 800, 000 km
   d. 850, 000 km

13. In this figure, if f_1 represents the parent body, which set of statements holds true?

   a. Area X < Area Y; the speed is greater for area X.
   b. Area X > Area Y; the speed is greater for area Y.
   c. Area X = Area Y; the speed is greater for area X.
   d. Area X = Area Y; the speed is greater for area Y.

7.2 Newton's Law of Universal Gravitation and Einstein's Theory of General Relativity

14. Rhea, with a radius of 7.63×10^5 m, is the second-largest moon of the planet Saturn. If the mass of Rhea is 2.31×10^{21} kg, what is the acceleration due to gravity on the surface of this moon?
   a. 2.65×10^{-1} m/s
   b. 2.02×10^5 m/s
   c. 2.65×10^{-1} m/s^2
   d. 2.02×10^5 m/s^2

15. Earth has a mass of 5.971×10^{24} kg and a radius of 6.371×10^6 m. Use the data to check the value of the gravitational constant.
   a. $6.66 \times 10^{-11}\ \frac{\text{N}\cdot\text{m}}{\text{kg}^2}$, it matches the value of the gravitational constant G.
   b. $1.05 \times 10^{-17}\ \frac{\text{N}\cdot\text{m}}{\text{kg}^2}$, it matches the value of the gravitational constant G.
   c. $6.66 \times 10^{-11}\ \frac{\text{N}\cdot\text{m}^2}{\text{kg}^2}$, it matches the value of the gravitational constant G.

d. $1.05 \times 10^{-17}\ \frac{\text{N} \cdot \text{m}^2}{\text{kg}^2}$, it matches the value of the gravitational constant G.

16. The orbit of the planet Mercury has a period of 88.0 days and an average radius of 5.791×10^{10} m. What is the mass of the sun?

a. 3.43×10^{19} kg
b. 1.99×10^{30} kg
c. 2.56×10^{29} kg
d. 1.48×10^{40} kg

Problems

7.1 Kepler's Laws of Planetary Motion

17. The closest Earth comes to the sun is 1.47×10^8 km, and Earth's farthest distance from the sun is 1.52×10^8 km. What is the area inside Earth's orbit?

a. 2.23×10^{16} km^2
b. 6.79×10^{16} km^2
c. 7.02×10^{16} km^2
d. 7.26×10^{16} km^2

18. Earth is 1.496×10^8 km from the sun, and Neptune is 4.490×10^9 km from the sun. What best represents the number of Earth years it takes for Neptune to complete one orbit around the sun?

a. 10 years
b. 30 years
c. 160 years
d. 900 years

Performance Task

7.2 Newton's Law of Universal Gravitation and Einstein's Theory of General Relativity

19. Design an experiment to test whether magnetic force is inversely proportional to the square of distance. Gravitational, magnetic, and electrical fields all act at a distance, but do they all follow the inverse square law? One difference in the forces related to these fields is that gravity is only attractive, but the other two can repel as well. In general, the inverse square law says that force F equals a constant C divided by the distance between objects, d, squared: $F = C/d^2$.
Incorporate these materials into your design:

- Two strong, permanent bar magnets
- A spring scale that can measure small forces
- A short ruler calibrated in millimeters

Use the magnets to study the relationship between attractive force and distance.

a. What will be the independent variable?
b. What will be the dependent variable?
c. How will you measure each of these variables?
d. If you plot the independent variable versus the dependent variable and the inverse square law is upheld, will the plot be a straight line? Explain.
e. Which plot would be a straight line if the inverse square law were upheld?

TEST PREP

Multiple Choice

7.1 Kepler's Laws of Planetary Motion

20. A planet of mass m circles a sun of mass M. Which distance changes throughout the planet's orbit?

a. $\overline{f_1 f_2}$
b. $\overline{mM}$
c. $\overline{Mf_2}$
d. $\overline{Mf_1}$

21. The focal point of the elliptical orbit of a moon is $50,000$ km from the center of the orbit. If the eccentricity of the orbit is 0.25, what is the length of the semi-major axis?

a. $12,500$ km
b. $100,000$ km
c. $200,000$ km
d. $400,000$ km

22. An artificial satellite orbits the Earth at a distance of 1.45×10^4 km from Earth's center. The moon orbits the Earth at a distance of 3.84×10^5 km once every 27.3 days. How long does it take the satellite to orbit the Earth?

a. 0.200 days
b. 3.07 days
c. 243 days
d. 3721 days

23. Earth is 1.496×10^8 km from the sun, and Venus is 1.08×10^8 km from the sun. One day on Venus is 243 Earth days long. What best represents the number of Venusian days in a Venusian year?

a. 0.78 days
b. 0.92 days
c. 1.08 days
d. 1.21 days

7.2 Newton's Law of Universal Gravitation and Einstein's Theory of General Relativity

24. What did the Cavendish experiment measure?
a. The mass of Earth
b. The gravitational constant
c. Acceleration due to gravity
d. The eccentricity of Earth's orbit

25. You have a mass of 55 kg and you have just landed on one of the moons of Jupiter where you have a weight of 67.9 N. What is the acceleration due to gravity, g, on the moon you are visiting?
a. $.810\text{m/s}^2$
b. 1.23m/s^2
c. 539m/s^2
d. 3735 m/s^2

26. A person is in an elevator that suddenly begins to descend. The person knows, intuitively, that the feeling of suddenly becoming lighter is because the elevator is accelerating downward. What other change would produce the same feeling? How does this demonstrate Einstein's postulate on which he based the theory of general relativity?
a. It would feel the same if the force of gravity suddenly became weaker. This illustrates Einstein's postulates that gravity and acceleration are indistinguishable.
b. It would feel the same if the force of gravity suddenly became stronger. This illustrates Einstein's postulates that gravity and acceleration are indistinguishable.
c. It would feel the same if the force of gravity suddenly became weaker. This illustrates Einstein's postulates that gravity and acceleration are distinguishable.
d. It would feel the same if the force of gravity suddenly became stronger. This illustrates Einstein's postulates that gravity and acceleration are distinguishable.

Short Answer

7.1 Kepler's Laws of Planetary Motion

27. Explain how the masses of a satellite and its parent body must compare in order to apply Kepler's laws of planetary motion.
a. The mass of the parent body must be much less than that of the satellite.
b. The mass of the parent body must be much greater than that of the satellite.
c. The mass of the parent body must be equal to the mass of the satellite.
d. There is no specific relationship between the masses for applying Kepler's laws of planetary motion.

28. Hyperion is a moon of the planet Saturn. Its orbit has an eccentricity of 0.123 and a semi-major axis of $1.48 \times 10^6\text{ km}$. How far is the center of the orbit from the center of Saturn?
a. $1.82 \times 10^5\text{ km}$
b. $3.64 \times 10^5\text{ km}$
c. $1.20 \times 10^7\text{ km}$
d. $2.41 \times 10^7\text{ km}$

29. The orbits of satellites are elliptical. Define an ellipse.
a. An ellipse is an open curve wherein the sum of the distance from the foci to any point on the curve is constant.
b. An ellipse is a closed curve wherein the sum of the distance from the foci to any point on the curve is constant.
c. An ellipse is an open curve wherein the distances from the two foci to any point on the curve are equal.
d. An ellipse is a closed curve wherein the distances from the two foci to any point on the curve are equal.

30. Mars has two moons, Deimos and Phobos. The orbit of Deimos has a period of 1.26 days and an average radius of $2.35 \times 10^3\text{ km}$. The average radius of the orbit of Phobos is $9.374 \times 10^3\text{ km}$. According to Kepler's third law of planetary motion, what is the period of Phobos?
a. 0.16 d
b. 0.50 d
c. 3.17 d
d. 10.0 d

7.2 Newton's Law of Universal Gravitation and Einstein's Theory of General Relativity

31. Newton's third law of motion says that, for every action force, there is a reaction force equal in magnitude but that acts in the opposite direction. Apply this law to gravitational forces acting between the Washington Monument and Earth.
a. The monument is attracted to Earth with a force equal to its weight, and Earth is attracted to the monument with a force equal to Earth's weight. The situation can be represented with two force vectors of unequal magnitude and pointing in the same direction.

b. The monument is attracted to Earth with a force equal to its weight, and Earth is attracted to the monument with a force equal to Earth's weight. The situation can be represented with two force vectors of unequal magnitude but pointing in opposite directions.
c. The monument is attracted to Earth with a force equal to its weight, and Earth is attracted to the monument with an equal force. The situation can be represented with two force vectors of equal magnitude and pointing in the same direction.
d. The monument is attracted to Earth with a force equal to its weight, and Earth is attracted to the monument with an equal force. The situation can be represented with two force vectors of equal magnitude but pointing in opposite directions.

32. True or false—Gravitational force is the attraction of the mass of one object to the mass of another. Light, either as a particle or a wave, has no rest mass. Despite this fact gravity bends a beam of light.
a. True
b. False

33. The average radius of Earth is 6.37×10^6 m. What is Earth's mass?
a. 9.35×10^{17} kg
b. 5.96×10^{24} kg
c. 3.79×10^{31} kg
d. 2.42×10^{38} kg

34. What is the gravitational force between two 60.0 kg people sitting 100 m apart?
a. 2.4×10^{-11} N
b. 2.4×10^{-9} N
c. 3.6×10^{-1} N
d. 3.6×10^{1} N

Extended Response

7.1 Kepler's Laws of Planetary Motion

35. The orbit of Halley"s Comet has an eccentricity of 0.967 and stretches to the edge of the solar system.
Part A. Describe the shape of the comet's orbit.
Part B. Compare the distance traveled per day when it is near the sun to the distance traveled per day when it is at the edge of the solar system.
Part C. Describe variations in the comet's speed as it completes an orbit. Explain the variations in terms of Kepler's second law of planetary motion.
a. Part A. The orbit is circular, with the sun at the center. Part B. The comet travels much farther when it is near the sun than when it is at the edge of the solar system. Part C. The comet decelerates as it approaches the sun and accelerates as it leaves the sun.
b. Part A. The orbit is circular, with the sun at the center. Part B. The comet travels much farther when it is near the sun than when it is at the edge of the solar system. Part C. The comet accelerates as it approaches the sun and decelerates as it leaves the sun.
c. Part A. The orbit is very elongated, with the sun near one end. Part B. The comet travels much farther when it is near the sun than when it is at the edge of the solar system. Part C. The comet decelerates as it approaches the sun and accelerates as it moves away from the sun.

36. For convenience, astronomers often use astronomical units (AU) to measure distances within the solar system. One AU equals the average distance from Earth to the sun. Halley's Comet returns once every 75.3 years. What is the average radius of the orbit of Halley's Comet in AU?
a. 0.002 AU
b. 0.056 AU
c. 17.8 AU
d. 653 AU

7.2 Newton's Law of Universal Gravitation and Einstein's Theory of General Relativity

37. It took scientists a long time to arrive at the understanding of gravity as explained by Galileo and Newton. They were hindered by two ideas that seemed like common sense but were serious misconceptions. First was the fact that heavier things fall faster than light things. Second, it was believed impossible that forces could act at a distance. Explain why these ideas persisted and why they prevented advances.
a. Heavier things fall faster than light things if they have less surface area and greater mass density. In the Renaissance and before, forces that acted at a distance were considered impossible, so people were skeptical about scientific theories that invoked such forces.
b. Heavier things fall faster than light things because they have greater surface area and less mass density. In the Renaissance and before, forces that act at a distance were considered impossible, so people were skeptical about scientific theories that invoked such forces.
c. Heavier things fall faster than light things because they have less surface area and greater mass density. In the Renaissance and before, forces that

act at a distance were considered impossible, so people were quick to accept scientific theories that invoked such forces.

d. Heavier things fall faster than light things because they have larger surface area and less mass density. In the Renaissance and before, forces that act at a distance were considered impossible because of people's faith in scientific theories.

38. The masses of Earth and the moon are 5.97×10^{24} kg and 7.35×10^{22} kg, respectively. The distance from Earth to the moon is 3.80×10^{5} km. At what point between the Earth and the moon are the opposing gravitational forces equal? (Use subscripts e and m to represent Earth and moon.)

a. 3.42×10^{5} km from the center of Earth
b. 3.80×10^{5} km from the center of Earth
c. 3.42×10^{6} km from the center of Earth
d. 3.10×10^{7} km from the center of Earth

CHAPTER 8
Momentum

Figure 8.1 NFC defensive backs Ronde Barber and Roy Williams along with linebacker Jeremiah Trotter gang tackle AFC running back LaDainian Tomlinson during the 2006 Pro Bowl in Hawaii. (United States Marine Corps)

Chapter Outline

8.1 Linear Momentum, Force, and Impulse

8.2 Conservation of Momentum

8.3 Elastic and Inelastic Collisions

INTRODUCTION We know from everyday use of the word *momentum* that it is a tendency to continue on course in the same direction. Newscasters speak of sports teams or politicians gaining, losing, or maintaining the momentum to win. As we learned when studying about inertia, which is Newton's first law of motion, every object or system has inertia—that is, a tendency for an object in motion to remain in motion or an object at rest to remain at rest. Mass is a useful variable that lets us quantify inertia. Momentum is mass in motion.

Momentum is important because it is conserved in isolated systems; this fact is convenient for solving problems where objects collide. The magnitude of momentum grows with greater mass and/or speed. For example, look at the football players in the photograph (Figure 8.1). They collide and fall to the ground. During their collisions, momentum will play a large part. In this chapter, we will learn about momentum, the different types of collisions, and how to use momentum equations to solve collision problems.

8.1 Linear Momentum, Force, and Impulse

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Describe momentum, what can change momentum, impulse, and the impulse-momentum theorem
- Describe Newton's second law in terms of momentum
- Solve problems using the impulse-momentum theorem

Section Key Terms

change in momentum	impulse	impulse–momentum theorem	linear momentum

Momentum, Impulse, and the Impulse-Momentum Theorem

Linear momentum is the product of a system's mass and its velocity. In equation form, linear momentum $\mathbf{p}$ is

$$\mathbf{p} = m\mathbf{v}.$$

You can see from the equation that momentum is directly proportional to the object's mass (m) and velocity ($\mathbf{v}$). Therefore, the greater an object's mass or the greater its velocity, the greater its momentum. A large, fast-moving object has greater momentum than a smaller, slower object.

Momentum is a vector and has the same direction as velocity $\mathbf{v}$. Since mass is a scalar, when velocity is in a negative direction (i.e., opposite the direction of motion), the momentum will also be in a negative direction; and when velocity is in a positive direction, momentum will likewise be in a positive direction. The SI unit for momentum is kg m/s.

Momentum is so important for understanding motion that it was called the *quantity of motion* by physicists such as Newton. Force influences momentum, and we can rearrange Newton's second law of motion to show the relationship between force and momentum.

Recall our study of Newton's second law of motion ($\mathbf{F}_{\text{net}} = m\mathbf{a}$). Newton actually stated his second law of motion in terms of momentum: The net external force equals the **change in momentum** of a system divided by the time over which it changes. The change in momentum is the difference between the final and initial values of momentum.

In equation form, this law is

$$\mathbf{F}_{\text{net}} = \frac{\Delta \mathbf{p}}{\Delta t},$$

where $\mathbf{F}_{\text{net}}$ is the net external force, $\Delta \mathbf{p}$ is the change in momentum, and Δt is the change in time.

We can solve for $\Delta \mathbf{p}$ by rearranging the equation

$$\mathbf{F}_{\text{net}} = \frac{\Delta \mathbf{p}}{\Delta t}$$

to be

$$\Delta \mathbf{p} = \mathbf{F}_{\text{net}} \Delta t.$$

$\mathbf{F}_{\text{net}} \Delta t$ is known as **impulse** and this equation is known as the **impulse-momentum theorem**. From the equation, we see that the impulse equals the average net external force multiplied by the time this force acts. It is equal to the change in momentum. *The effect of a force on an object depends on how long it acts, as well as the strength of the force.* Impulse is a useful concept because it quantifies the effect of a force. A very large force acting for a short time can have a great effect on the momentum of an object, such as the force of a racket hitting a tennis ball. A small force could cause the same change in momentum, but it would have to act for a much longer time.

Newton's Second Law in Terms of Momentum

When Newton's second law is expressed in terms of momentum, it can be used for solving problems where mass varies, since $\Delta \mathbf{p} = \Delta(m\mathbf{v})$. In the more traditional form of the law that you are used to working with, mass is assumed to be constant. In fact, this traditional form is a special case of the law, where mass is constant. $\mathbf{F}_{\text{net}} = m\mathbf{a}$ is actually derived from the equation:

$$\mathbf{F}_{\text{net}} = \frac{\Delta \mathbf{p}}{\Delta t}$$

For the sake of understanding the relationship between Newton's second law in its two forms, let's recreate the derivation of $\mathbf{F}_{\text{net}} = m\mathbf{a}$ from

$$\mathbf{F}_{\text{net}} = \frac{\Delta \mathbf{p}}{\Delta t}$$

by substituting the definitions of acceleration and momentum.

The change in momentum $\Delta \mathbf{p}$ is given by

$$\Delta \mathbf{p} = \Delta(m\mathbf{v}).$$

If the mass of the system is constant, then

$$\Delta(m\mathbf{v}) = m\Delta\mathbf{v}.$$

By substituting $m\Delta\mathbf{v}$ for $\Delta\mathbf{p}$, Newton's second law of motion becomes

$$\mathbf{F}_{\text{net}} = \frac{\Delta \mathbf{p}}{\Delta t} = \frac{m\Delta\mathbf{v}}{\Delta t}$$

for a constant mass.

Because

$$\frac{\Delta \mathbf{v}}{\Delta t} = \mathbf{a},$$

we can substitute to get the familiar equation

$$\mathbf{F}_{\text{net}} = m\mathbf{a}$$

when the mass of the system is constant.

TIPS FOR SUCCESS

We just showed how $\mathbf{F}_{\text{net}} = m\mathbf{a}$ applies only when the mass of the system is constant. An example of when this formula would not apply would be a moving rocket that burns enough fuel to significantly change the mass of the rocket. In this case, you would need to use Newton's second law expressed in terms of momentum to account for the changing mass.

Snap Lab

Hand Movement and Impulse

In this activity you will experiment with different types of hand motions to gain an intuitive understanding of the relationship between force, time, and impulse.

- one ball
- one tub filled with water

Procedure:

1. Try catching a ball while *giving* with the ball, pulling your hands toward your body.
2. Next, try catching a ball while keeping your hands still.
3. Hit water in a tub with your full palm. Your full palm represents a swimmer doing a belly flop.
4. After the water has settled, hit the water again by diving your hand with your fingers first into the water. Your diving hand represents a swimmer doing a dive.
5. Explain what happens in each case and why.

GRASP CHECK

What are some other examples of motions that impulse affects?

a. a football player colliding with another, or a car moving at a constant velocity
b. a car moving at a constant velocity, or an object moving in the projectile motion
c. a car moving at a constant velocity, or a racket hitting a ball
d. a football player colliding with another, or a racket hitting a ball

LINKS TO PHYSICS

Engineering: Saving Lives Using the Concept of Impulse

Cars during the past several decades have gotten much safer. Seat belts play a major role in automobile safety by preventing people from flying into the windshield in the event of a crash. Other safety features, such as airbags, are less visible or obvious, but are also effective at making auto crashes less deadly (see Figure 8.2). Many of these safety features make use of the concept of impulse from physics. Recall that impulse is the net force multiplied by the duration of time of the impact. This was expressed mathematically as $\Delta \mathbf{p} = \mathbf{F}_{\text{net}} \Delta t$.

Figure 8.2 Vehicles have safety features like airbags and seat belts installed.

Airbags allow the net force on the occupants in the car to act over a much longer time when there is a sudden stop. The momentum change is the same for an occupant whether an airbag is deployed or not. But the force that brings the occupant to a stop will be much less if it acts over a larger time. By rearranging the equation for impulse to solve for force $\mathbf{F}_{\text{net}} = \frac{\Delta \mathbf{p}}{\Delta t}$, you can see how increasing Δt while $\Delta \mathbf{p}$ stays the same will decrease $\mathbf{F}_{\text{net}}$. This is another example of an inverse relationship. Similarly, a padded dashboard increases the time over which the force of impact acts, thereby reducing the force of impact.

Cars today have many plastic components. One advantage of plastics is their lighter weight, which results in better gas mileage. Another advantage is that a car will crumple in a collision, especially in the event of a head-on collision. A longer collision time means the force on the occupants of the car will be less. Deaths during car races decreased dramatically when the rigid frames of racing cars were replaced with parts that could crumple or collapse in the event of an accident.

GRASP CHECK

You may have heard the advice to bend your knees when jumping. In this example, a friend dares you to jump off of a park bench onto the ground without bending your knees. You, of course, refuse. Explain to your friend why this would be a foolish thing. Show it using the impulse-momentum theorem.

a. Bending your knees increases the time of the impact, thus decreasing the force.
b. Bending your knees decreases the time of the impact, thus decreasing the force.
c. Bending your knees increases the time of the impact, thus increasing the force.
d. Bending your knees decreases the time of the impact, thus increasing the force.

Solving Problems Using the Impulse-Momentum Theorem

WORKED EXAMPLE

Calculating Momentum: A Football Player and a Football

(a) Calculate the momentum of a 110 kg football player running at 8 m/s. (b) Compare the player's momentum with the momentum of a 0.410 kg football thrown hard at a speed of 25 m/s.

Strategy

No information is given about the direction of the football player or the football, so we can calculate only the magnitude of the momentum, *p*. (A symbol in italics represents magnitude.) In both parts of this example, the magnitude of momentum can be calculated directly from the definition of momentum:

$$\mathbf{p} = m\mathbf{v}$$

Solution for (a)

To find the player's momentum, substitute the known values for the player's mass and speed into the equation.

$$\mathbf{p}_{\text{player}} = (110\ \text{kg})(8\ \text{m/s}) = 880\ \text{kg} \cdot \text{m/s}$$

Solution for (b)

To find the ball's momentum, substitute the known values for the ball's mass and speed into the equation.

$$\mathbf{p}_{\text{ball}} = (0.410\ \text{kg})(25\ \text{m/s}) = 10.25\ \text{kg} \cdot \text{m/s}$$

The ratio of the player's momentum to the ball's momentum is

$$\frac{\mathbf{p}_{\text{player}}}{\mathbf{p}_{\text{ball}}} = \frac{880}{10.3} = 85.9\ .$$

Discussion

Although the ball has greater velocity, the player has a much greater mass. Therefore, the momentum of the player is about 86 times greater than the momentum of the football.

WORKED EXAMPLE

Calculating Force: Venus Williams' Racquet

During the 2007 French Open, Venus Williams (Figure 8.3) hit the fastest recorded serve in a premier women's match, reaching a speed of 58 m/s (209 km/h). What was the average force exerted on the 0.057 kg tennis ball by Williams' racquet? Assume that the ball's speed just after impact was 58 m/s, the horizontal velocity before impact is negligible, and that the ball remained in contact with the racquet for 5 ms (milliseconds).

Figure 8.3 Venus Williams playing in the 2013 US Open (Edwin Martinez, Flickr)

Strategy

Recall that Newton's second law stated in terms of momentum is

$$\mathbf{F}_{\text{net}} = \frac{\Delta \mathbf{p}}{\Delta t}.$$

As noted above, when mass is constant, the change in momentum is given by

$$\Delta \mathbf{p} = m\Delta \mathbf{v} = m(\mathbf{v}_{\text{f}} - \mathbf{v}_{\text{i}}),$$

where $\mathbf{v}_{\text{f}}$ is the final velocity and $\mathbf{v}_{\text{i}}$ is the initial velocity. In this example, the velocity just after impact and the change in time are given, so after we solve for $\Delta \mathbf{p}$, we can use $\mathbf{F}_{\text{net}} = \frac{\Delta \mathbf{p}}{\Delta t}$ to find the force.

Solution

To determine the change in momentum, substitute the values for mass and the initial and final velocities into the equation above.

$$\begin{aligned} \Delta \mathbf{p} &= m(\mathbf{v}_f - \mathbf{v}_i) \\ &= (0.057\text{ kg})\,(58\text{ m/s} - 0\text{ m/s}) \\ &= 3.306\text{ kg·m/s} \approx 3.3\text{ kg·m/s} \end{aligned} \tag{8.1}$$

Now we can find the magnitude of the net external force using $\mathbf{F}_{\text{net}} = \frac{\Delta \mathbf{p}}{\Delta t}$

$$\begin{aligned} \mathbf{F}_{\text{net}} &= \frac{\Delta \mathbf{p}}{\Delta t} = \frac{3.306}{5\times10^{-3}} \\ &= 661\text{ N} \\ &\approx 660\text{ N}. \end{aligned} \tag{8.2}$$

Discussion

This quantity was the average force exerted by Venus Williams' racquet on the tennis ball during its brief impact. This problem could also be solved by first finding the acceleration and then using $\mathbf{F}_{\text{net}} = m\mathbf{a}$, but we would have had to do one more step. In this case, using momentum was a shortcut.

Practice Problems

1. What is the momentum of a bowling ball with mass 5 kg and velocity 10 m/s?
   a. $0.5\text{ kg} \cdot \text{m/s}$
   b. $2\text{ kg} \cdot \text{m/s}$
   c. $15\text{ kg} \cdot \text{m/s}$
   d. $50\text{ kg} \cdot \text{m/s}$

2. What will be the change in momentum caused by a net force of 120 N acting on an object for 2 seconds?
   a. $60\text{ kg} \cdot \text{m/s}$
   b. $118\text{ kg} \cdot \text{m/s}$
   c. $122\text{ kg} \cdot \text{m/s}$
   d. $240\text{ kg} \cdot \text{m/s}$

Check Your Understanding

3. What is linear momentum?
   a. the sum of a system's mass and its velocity
   b. the ratio of a system's mass to its velocity
   c. the product of a system's mass and its velocity
   d. the product of a system's moment of inertia and its velocity

4. If an object's mass is constant, what is its momentum proportional to?
   a. Its velocity
   b. Its weight
   c. Its displacement
   d. Its moment of inertia

5. What is the equation for Newton's second law of motion, in terms of mass, velocity, and time, when the mass of the system is

constant?

a. $\mathbf{F}_{\text{net}} = \frac{\Delta \mathbf{v}}{\Delta m \Delta t}$
b. $\mathbf{F}_{\text{net}} = \frac{m\Delta t}{\Delta \mathbf{v}}$
c. $\mathbf{F}_{\text{net}} = \frac{m\Delta \mathbf{v}}{\Delta t}$
d. $\mathbf{F}_{\text{net}} = \frac{\Delta m \Delta \mathbf{v}}{\Delta t}$

6. Give an example of a system whose mass is not constant.

a. A spinning top
b. A baseball flying through the air
c. A rocket launched from Earth
d. A block sliding on a frictionless inclined plane

8.2 Conservation of Momentum

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Describe the law of conservation of momentum verbally and mathematically

Section Key Terms

angular momentum	isolated system	law of conservation of momentum

Conservation of Momentum

It is important we realize that momentum is conserved during collisions, explosions, and other events involving objects in motion. To say that a quantity is conserved means that it is constant throughout the event. In the case of conservation of momentum, the total momentum in the system remains the same before and after the collision.

You may have noticed that momentum was *not* conserved in some of the examples previously presented in this chapter. where forces acting on the objects produced large changes in momentum. Why is this? The systems of interest considered in those problems were not inclusive enough. If the systems were expanded to include more objects, then momentum would in fact be conserved in those sample problems. It is always possible to find a larger system where momentum is conserved, even though momentum changes for individual objects within the system.

For example, if a football player runs into the goalpost in the end zone, a force will cause him to bounce backward. His momentum is obviously greatly changed, and considering only the football player, we would find that momentum is not conserved. However, the system can be expanded to contain the entire Earth. Surprisingly, Earth also recoils—conserving momentum—because of the force applied to it through the goalpost. The effect on Earth is not noticeable because it is so much more massive than the player, but the effect is real.

Next, consider what happens if the masses of two colliding objects are more similar than the masses of a football player and Earth—in the example shown in Figure 8.4 of one car bumping into another. Both cars are coasting in the same direction when the lead car, labeled m_2, is bumped by the trailing car, labeled m_1. The only unbalanced force on each car is the force of the collision, assuming that the effects due to friction are negligible. Car $m1$ slows down as a result of the collision, losing some momentum, while car $m2$ speeds up and gains some momentum. If we choose the system to include both cars and assume that friction is negligible, then the momentum of the two-car system should remain constant. Now we will prove that the total momentum of the two-car system does in fact remain constant, and is therefore conserved.

Figure 8.4 Car of mass m_1 moving with a velocity of $\mathbf{v}_1$ bumps into another car of mass m_2 and velocity $\mathbf{v}_2$. As a result, the first car slows down to a velocity of $\mathbf{v}'_1$ and the second speeds up to a velocity of $\mathbf{v}'_2$. The momentum of each car is changed, but the total momentum $\mathbf{p}_{tot}$ of the two cars is the same before and after the collision if you assume friction is negligible.

Using the impulse-momentum theorem, the change in momentum of car 1 is given by

$$\Delta\mathbf{p}_1 = \mathbf{F}_1 \Delta t,$$

where $\mathbf{F}_1$ is the force on car 1 due to car 2, and Δt is the time the force acts, or the duration of the collision.

Similarly, the change in momentum of car 2 is $\Delta\mathbf{p}_2 = \mathbf{F}_2 \Delta t$ where $\mathbf{F}_2$ is the force on car 2 due to car 1, and we assume the duration of the collision Δt is the same for both cars. We know from Newton's third law of motion that $\mathbf{F}_2 = -\mathbf{F}_1$, and so $\Delta\mathbf{p}_2 = -\mathbf{F}_1 \Delta t = -\Delta\mathbf{p}_1$.

Therefore, the changes in momentum are equal and opposite, and $\Delta\mathbf{p}_1 + \Delta\mathbf{p}_2 = 0$.

Because the changes in momentum add to zero, the total momentum of the two-car system is constant. That is,

$$\mathbf{p}_1 + \mathbf{p}_2 = \text{constant}$$

$$\mathbf{p}_1 + \mathbf{p}_2 = \mathbf{p}'_1 + \mathbf{p}'_2,$$

where $\mathbf{p}'_1$ and $\mathbf{p}'_2$ are the momenta of cars 1 and 2 after the collision.

This result that momentum is conserved is true not only for this example involving the two cars, but for any system where the net external force is zero, which is known as an **isolated system**. The **law of conservation of momentum** states that for an isolated system with any number of objects in it, the total momentum is conserved. In equation form, the law of conservation of momentum for an isolated system is written as

$$\mathbf{p}_{tot} = \text{constant}$$

or

$$\mathbf{p}_{tot} = \mathbf{p}'_{tot},$$

where $\mathbf{p}_{tot}$ is the total momentum, or the sum of the momenta of the individual objects in the system at a given time, and $\mathbf{p}'_{tot}$ is the total momentum some time later.

The conservation of momentum principle can be applied to systems as diverse as a comet striking the Earth or a gas containing huge numbers of atoms and molecules. Conservation of momentum appears to be violated only when the net external force is not zero. But another larger system can always be considered in which momentum is conserved by simply including the source of the external force. For example, in the collision of two cars considered above, the two-car system conserves momentum while each one-car system does not.

TIPS FOR SUCCESS

Momenta is the plural form of the word momentum. One object is said to have momentum, but two or more objects are said to have *momenta*.

FUN IN PHYSICS

Angular Momentum in Figure Skating

So far we have covered linear momentum, which describes the inertia of objects traveling in a straight line. But we know that many objects in nature have a curved or circular path. Just as linear motion has linear momentum to describe its tendency to move forward, circular motion has the equivalent **angular momentum** to describe how rotational motion is carried forward.

This is similar to how torque is analogous to force, angular acceleration is analogous to translational acceleration, and mr^2 is analogous to mass or inertia. You may recall learning that the quantity mr^2 is called the rotational inertia or moment of inertia of a point mass m at a distance r from the center of rotation.

We already know the equation for linear momentum, $\mathbf{p} = m\mathbf{v}$. Since angular momentum is analogous to linear momentum, the moment of inertia (I) is analogous to mass, and angular velocity ($\boldsymbol{\omega}$) is analogous to linear velocity, it makes sense that angular momentum ($\mathbf{L}$) is defined as

$$\mathbf{L} = I\boldsymbol{\omega}$$

Angular momentum is conserved when the net external torque ($\boldsymbol{\tau}$) is zero, just as linear momentum is conserved when the net external force is zero.

Figure skaters take advantage of the conservation of angular momentum, likely without even realizing it. In Figure 8.5, a figure skater is executing a spin. The net torque on her is very close to zero, because there is relatively little friction between her skates and the ice, and because the friction is exerted very close to the pivot point. Both $\mathbf{F}$ and r are small, and so $\boldsymbol{\tau}$ is negligibly small.

Figure 8.5 (a) An ice skater is spinning on the tip of her skate with her arms extended. In the next image, (b), her rate of spin increases greatly when she pulls in her arms.

Consequently, she can spin for quite some time. She can do something else, too. She can increase her rate of spin by pulling her arms and legs in. Why does pulling her arms and legs in increase her rate of spin? The answer is that her angular momentum is constant, so that $\mathbf{L} = \mathbf{L}'$.

Expressing this equation in terms of the moment of inertia,

$$I\boldsymbol{\omega} = I'\boldsymbol{\omega}',$$

where the primed quantities refer to conditions after she has pulled in her arms and reduced her moment of inertia. Because I' is smaller, the angular velocity $\boldsymbol{\omega}'$ must increase to keep the angular momentum constant. This allows her to spin much faster without exerting any extra torque.

A video (http://openstax.org/l/28figureskater) is also available that shows a real figure skater executing a spin. It discusses the physics of spins in figure skating.

GRASP CHECK

Based on the equation $\mathbf{L} = I\omega$, how would you expect the moment of inertia of an object to affect angular momentum? How would angular velocity affect angular momentum?

a. Large moment of inertia implies large angular momentum, and large angular velocity implies large angular momentum.
b. Large moment of inertia implies small angular momentum, and large angular velocity implies small angular momentum.
c. Large moment of inertia implies large angular momentum, and large angular velocity implies small angular momentum.
d. Large moment of inertia implies small angular momentum, and large angular velocity implies large angular momentum.

Check Your Understanding

7. When is momentum said to be conserved?
   a. When momentum is changing during an event
   b. When momentum is increasing during an event
   c. When momentum is decreasing during an event
   d. When momentum is constant throughout an event

8. A ball is hit by a racket and its momentum changes. How is momentum conserved in this case?
   a. Momentum of the system can never be conserved in this case.
   b. Momentum of the system is conserved if the momentum of the racket is not considered.
   c. Momentum of the system is conserved if the momentum of the racket is also considered.
   d. Momentum of the system is conserved if the momenta of the racket and the player are also considered.

9. State the law of conservation of momentum.
   a. Momentum is conserved for an isolated system with any number of objects in it.
   b. Momentum is conserved for an isolated system with an even number of objects in it.
   c. Momentum is conserved for an interacting system with any number of objects in it.
   d. Momentum is conserved for an interacting system with an even number of objects in it.

8.3 Elastic and Inelastic Collisions

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Distinguish between elastic and inelastic collisions
- Solve collision problems by applying the law of conservation of momentum

Section Key Terms

elastic collision	inelastic collision	point masses	recoil

Elastic and Inelastic Collisions

When objects collide, they can either stick together or bounce off one another, remaining separate. In this section, we'll cover these two different types of collisions, first in one dimension and then in two dimensions.

In an **elastic collision**, the objects separate after impact and don't lose any of their kinetic energy. Kinetic energy is the energy of motion and is covered in detail elsewhere. The law of conservation of momentum is very useful here, and it can be used whenever the net external force on a system is zero. Figure 8.6 shows an elastic collision where momentum is conserved.

Figure 8.6 The diagram shows a one-dimensional elastic collision between two objects.

An animation of an elastic collision between balls can be seen by watching this video (http://openstax.org/l/28elasticball) . It replicates the elastic collisions between balls of varying masses.

Perfectly elastic collisions can happen only with subatomic particles. Everyday observable examples of perfectly elastic collisions don't exist—some kinetic energy is always lost, as it is converted into heat transfer due to friction. However, collisions between everyday objects are almost perfectly elastic when they occur with objects and surfaces that are nearly frictionless, such as with two steel blocks on ice.

Now, to solve problems involving one-dimensional elastic collisions between two objects, we can use the equation for conservation of momentum. First, the equation for conservation of momentum for two objects in a one-dimensional collision is

$$\mathbf{p}_1 + \mathbf{p}_2 = \mathbf{p}'_1 + \mathbf{p}'_2 (\mathbf{F}_{\text{net}} = 0).$$

Substituting the definition of momentum $\mathbf{p} = m\mathbf{v}$ for each initial and final momentum, we get

$$m_1\mathbf{v}_1 + m_2\mathbf{v}_2 = m_1\mathbf{v}'_1 + m_2\mathbf{v}'_2,$$

where the primes (') indicate values after the collision; In some texts, you may see *i* for initial (before collision) and *f* for final (after collision). The equation assumes that the mass of each object does not change during the collision.

WATCH PHYSICS

Momentum: Ice Skater Throws a Ball

This video covers an elastic collision problem in which we find the ***recoil*** *velocity* of an ice skater who throws a ball straight forward. To clarify, Sal is using the equation

$$m_{\text{ball}}\mathbf{V}_{\text{ball}} + m_{\text{skater}}\mathbf{V}_{\text{skater}} = m_{\text{ball}}\mathbf{v}'_{\text{ball}} + m_{\text{skater}}\mathbf{v}'_{\text{skater}}.$$

Click to view content (https://www.khanacademy.org/embed_video?v=vPkkCOlGND4)

GRASP CHECK

The resultant vector of the addition of vectors $\vec{a}$ and $\vec{b}$ is $\vec{r}$. The magnitudes of $\vec{a}$, $\vec{b}$, and $\vec{r}$ are A, B, and R, respectively. Which of the following is true?

a. $R_x + R_y = 0$

b. $A_x + A_y = \vec{A}$

c. $A_x + B_y = B_x + A_y$
d. $A_x + B_x = R_x$

Now, let us turn to the second type of collision. An **inelastic collision** is one in which objects stick together after impact, and kinetic energy is *not* conserved. This lack of conservation means that the forces between colliding objects may convert kinetic energy to other forms of energy, such as potential energy or thermal energy. The concepts of energy are discussed more thoroughly elsewhere. For inelastic collisions, kinetic energy may be lost in the form of heat. Figure 8.7 shows an example of an inelastic collision. Two objects that have equal masses head toward each other at equal speeds and then stick together. The two objects come to rest after sticking together, conserving momentum but not kinetic energy after they collide. Some of the energy of motion gets converted to thermal energy, or heat.

Figure 8.7 A one-dimensional inelastic collision between two objects. Momentum is conserved, but kinetic energy is not conserved. (a) Two objects of equal mass initially head directly toward each other at the same speed. (b) The objects stick together, creating a perfectly inelastic collision. In the case shown in this figure, the combined objects stop; This is not true for all inelastic collisions.

Since the two objects stick together after colliding, they move together at the same speed. This lets us simplify the conservation of momentum equation from

$$m_1\mathbf{v}_1 + m_2\mathbf{v}_2 = m_1\mathbf{v}'_1 + m_2\mathbf{v}'_2$$

to

$$m_1\mathbf{v}_1 + m_2\mathbf{v}_2 = (m_1 + m_2)\mathbf{v}'$$

for inelastic collisions, where $\mathbf{v}'$ is the final velocity for both objects as they are stuck together, either in motion or at rest.

WATCH PHYSICS

Introduction to Momentum

This video reviews the definitions of momentum and impulse. It also covers an example of using conservation of momentum to solve a problem involving an inelastic collision between a car with constant velocity and a stationary truck. Note that Sal accidentally gives the unit for impulse as Joules; it is actually N · s or k · gm/s.

Click to view content (https://www.khanacademy.org/embed_video?v=XFhntPxowoU)

GRASP CHECK

How would the final velocity of the car-plus-truck system change if the truck had some initial velocity moving in the same direction as the car? What if the truck were moving in the opposite direction of the car initially? Why?

a. If the truck was initially moving in the same direction as the car, the final velocity would be greater. If the truck was initially moving in the opposite direction of the car, the final velocity would be smaller.
b. If the truck was initially moving in the same direction as the car, the final velocity would be smaller. If the truck was initially moving in the opposite direction of the car, the final velocity would be greater.
c. The direction in which the truck was initially moving would not matter. If the truck was initially moving in either

direction, the final velocity would be smaller.

d. The direction in which the truck was initially moving would not matter. If the truck was initially moving in either direction, the final velocity would be greater.

Snap Lab

Ice Cubes and Elastic Collisions

In this activity, you will observe an elastic collision by sliding an ice cube into another ice cube on a smooth surface, so that a negligible amount of energy is converted to heat.

- Several ice cubes (The ice must be in the form of cubes.)
- A smooth surface

Procedure

1. Find a few ice cubes that are about the same size and a smooth kitchen tabletop or a table with a glass top.
2. Place the ice cubes on the surface several centimeters away from each other.
3. Flick one ice cube toward a stationary ice cube and observe the path and velocities of the ice cubes after the collision. Try to avoid edge-on collisions and collisions with rotating ice cubes.
4. Explain the speeds and directions of the ice cubes using momentum.

GRASP CHECK

Was the collision elastic or inelastic?

a. perfectly elastic
b. perfectly inelastic
c. Nearly perfect elastic
d. Nearly perfect inelastic

TIPS FOR SUCCESS

Here's a trick for remembering which collisions are elastic and which are inelastic: Elastic is a bouncy material, so when objects *bounce* off one another in the collision and separate, it is an elastic collision. When they don't, the collision is inelastic.

Solving Collision Problems

The Khan Academy videos referenced in this section show examples of elastic and inelastic collisions in one dimension. In one-dimensional collisions, the incoming and outgoing velocities are all along the same line. But what about collisions, such as those between billiard balls, in which objects scatter to the side? These are two-dimensional collisions, and just as we did with two-dimensional forces, we will solve these problems by first choosing a coordinate system and separating the motion into its x and y components.

One complication with two-dimensional collisions is that the objects might rotate before or after their collision. For example, if two ice skaters hook arms as they pass each other, they will spin in circles. We will not consider such rotation until later, and so for now, we arrange things so that no rotation is possible. To avoid rotation, we consider only the scattering of **point masses**—that is, structureless particles that cannot rotate or spin.

We start by assuming that $\mathbf{F}_{\text{net}} = 0$, so that momentum $\mathbf{p}$ is conserved. The simplest collision is one in which one of the particles is initially at rest. The best choice for a coordinate system is one with an axis parallel to the velocity of the incoming particle, as shown in Figure 8.8. Because momentum is conserved, the components of momentum along the x- and y-axes, displayed as $\mathbf{p}_x$ and $\mathbf{p}_y$, will also be conserved. With the chosen coordinate system, $\mathbf{p}_y$ is initially zero and $\mathbf{p}_x$ is the momentum of the incoming particle.

Figure 8.8 A two-dimensional collision with the coordinate system chosen so that m_2 is initially at rest and $\mathbf{v}_1$ is parallel to the x-axis.

Now, we will take the conservation of momentum equation, $\mathbf{p}_1 + \mathbf{p}_2 = \mathbf{p}'_1 + \mathbf{p}'_2$ and break it into its x and y components.

Along the x-axis, the equation for conservation of momentum is

$$\mathbf{p}_{1x} + \mathbf{p}_{2x} = \mathbf{p}'_{1x} + \mathbf{p}'_{2x}.$$

In terms of masses and velocities, this equation is

$$m_1\mathbf{v}_{1\ x} + m_2\mathbf{v}_{2\ x} = m_1\mathbf{v}'_{1\ x} + m_2\mathbf{v}'_{2\ x}. \tag{8.3}$$

But because particle 2 is initially at rest, this equation becomes

$$m_1\mathbf{v}_{1\ x} = m_1\mathbf{v}'_{1\ x} + m_2\mathbf{v}'_{2\ x}. \tag{8.4}$$

The components of the velocities along the x-axis have the form $\mathbf{v}\cos\theta$. Because particle 1 initially moves along the x-axis, we find $\mathbf{v}_{1x} = \mathbf{v}_1$. Conservation of momentum along the x-axis gives the equation

$$m_1\mathbf{v}_1 = m_1\mathbf{v}'_1\cos\theta_1 + m_2\mathbf{v}'_2\cos\theta_2,$$

where θ_1 and θ_2 are as shown in Figure 8.8.

Along the y-axis, the equation for conservation of momentum is

$$\mathbf{p}_{1\ y} + \mathbf{p}_{2\ y} = \mathbf{p}'_{1\ y} + \mathbf{p}'_{2\ y}, \tag{8.5}$$

or

$$m_1\mathbf{v}_{1\ y} + m_2\mathbf{v}_{2\ y} = m_1\mathbf{v}'_{1\ y} + m_2\mathbf{v}'_{2\ y}. \tag{8.6}$$

But $\mathbf{v}_1y$ is zero, because particle 1 initially moves along the x-axis. Because particle 2 is initially at rest, $\mathbf{v}_2y$ is also zero. The equation for conservation of momentum along the y-axis becomes

$$0 = m_1\mathbf{v}'_1y + m_2\mathbf{v}'_2y. \tag{8.7}$$

The components of the velocities along the y-axis have the form $\mathbf{v}\sin\theta$. Therefore, conservation of momentum along the y-axis gives the following equation:

$$0 = m_1\mathbf{v}'_1\sin\theta_1 + m_2\mathbf{v}'_2\sin\theta_2$$

Virtual Physics

Collision Lab

In this simulation, you will investigate collisions on an air hockey table. Place checkmarks next to the momentum vectors

and momenta diagram options. Experiment with changing the masses of the balls and the initial speed of ball 1. How does this affect the momentum of each ball? What about the total momentum? Next, experiment with changing the elasticity of the collision. You will notice that collisions have varying degrees of elasticity, ranging from perfectly elastic to perfectly inelastic.

Click to view content (https://archive.cnx.org/specials/2c7acb3c-2fbd-11e5-b2d9-e7f92291703c/collision-lab/)

GRASP CHECK

If you wanted to maximize the velocity of ball 2 after impact, how would you change the settings for the masses of the balls, the initial speed of ball 1, and the elasticity setting? Why? Hint—Placing a checkmark next to the velocity vectors and removing the momentum vectors will help you visualize the velocity of ball 2, and pressing the More Data button will let you take readings.

a. Maximize the mass of ball 1 and initial speed of ball 1; minimize the mass of ball 2; and set elasticity to 50 percent.
b. Maximize the mass of ball 2 and initial speed of ball 1; minimize the mass of ball 1; and set elasticity to 100 percent.
c. Maximize the mass of ball 1 and initial speed of ball 1; minimize the mass of ball 2; and set elasticity to 100 percent.
d. Maximize the mass of ball 2 and initial speed of ball 1; minimize the mass of ball 1; and set elasticity to 50 percent.

WORKED EXAMPLE

Calculating Velocity: Inelastic Collision of a Puck and a Goalie

Find the recoil velocity of a 70 kg ice hockey goalie who catches a 0.150-kg hockey puck slapped at him at a velocity of 35 m/s. Assume that the goalie is at rest before catching the puck, and friction between the ice and the puck-goalie system is negligible (see Figure 8.9).

Figure 8.9 An ice hockey goalie catches a hockey puck and recoils backward in an inelastic collision.

Strategy

Momentum is conserved because the net external force on the puck-goalie system is zero. Therefore, we can use conservation of momentum to find the final velocity of the puck and goalie system. Note that the initial velocity of the goalie is zero and that the final velocity of the puck and goalie are the same.

Solution

For an inelastic collision, conservation of momentum is

$$m_1\mathbf{v}_1 + m_2\mathbf{v}_2 = (m_1 + m_2)\mathbf{v}', \tag{8.8}$$

where $\mathbf{v}'$ is the velocity of both the goalie and the puck after impact. Because the goalie is initially at rest, we know $\mathbf{v}_2 = 0$. This simplifies the equation to

$$m_1\mathbf{v}_1 = (m_1 + m_2)\mathbf{v}'. \tag{8.9}$$

Solving for $\mathbf{v}'$ yields

$$\mathbf{v}' = \left(\frac{m_1}{m_1 + m_2}\right)\mathbf{v}_1. \tag{8.10}$$

Entering known values in this equation, we get

$$\begin{aligned}\mathbf{v}' &= \left(\frac{0.150\text{ kg}}{70.0\text{ kg}+0.150\text{ kg}}\right)(35\text{ m/s}) \\ &= 7.48 \times 10^{-2}\text{m/s}.\end{aligned} \tag{8.11}$$

Discussion

This recoil velocity is small and in the same direction as the puck's original velocity.

WORKED EXAMPLE

Calculating Final Velocity: Elastic Collision of Two Carts

Two hard, steel carts collide head-on and then ricochet off each other in opposite directions on a frictionless surface (see Figure 8.10). Cart 1 has a mass of 0.350 kg and an initial velocity of 2 m/s. Cart 2 has a mass of 0.500 kg and an initial velocity of –0.500 m/s. After the collision, cart 1 recoils with a velocity of –4 m/s. What is the final velocity of cart 2?

Figure 8.10 Two carts collide with each other in an elastic collision.

Strategy

Since the track is frictionless, $\mathbf{F}_{\text{net}} = 0$ and we can use conservation of momentum to find the final velocity of cart 2.

Solution

As before, the equation for conservation of momentum for a one-dimensional elastic collision in a two-object system is

$$m_1\mathbf{v}_1 + m_2\mathbf{v}_2 = m_1\mathbf{v}'_1 + m_2\mathbf{v}'_2. \tag{8.12}$$

The only unknown in this equation is $\mathbf{v}'_2$. Solving for $\mathbf{v}'_2$ and substituting known values into the previous equation yields

$$\begin{aligned}\mathbf{v}'_2 &= \frac{m_1\mathbf{v}_1 + m_2\mathbf{v}_2 - m_1\mathbf{v}'_1}{m_2} \\ &= \frac{(0.350\text{kg})(2.00\text{m/s}) + (0.500\text{ kg})(-0.500\text{ m/s}) - (0.350\text{kg})(-4.00\text{m/s})}{0.500\text{kg}} \\ &= 3.70\text{m/s}.\end{aligned} \tag{8.13}$$

Discussion

The final velocity of cart 2 is large and positive, meaning that it is moving to the right after the collision.

WORKED EXAMPLE

Calculating Final Velocity in a Two-Dimensional Collision

Suppose the following experiment is performed (Figure 8.11). An object of mass 0.250 kg (m_1) is slid on a frictionless surface into a dark room, where it strikes an initially stationary object of mass 0.400 kg (m_2). The 0.250 kg object emerges from the room at an angle of 45° with its incoming direction. The speed of the 0.250 kg object is originally 2 m/s and is 1.50 m/s after the collision. Calculate the magnitude and direction of the velocity (v'_2 and θ_2) of the 0.400 kg object after the collision.

Figure 8.11 The incoming object of mass m_1 is scattered by an initially stationary object. Only the stationary object's mass m_2 is known. By measuring the angle and speed at which the object of mass m_1 emerges from the room, it is possible to calculate the magnitude and direction of the initially stationary object's velocity after the collision.

Strategy

Momentum is conserved because the surface is frictionless. We chose the coordinate system so that the initial velocity is parallel to the x-axis, and conservation of momentum along the x- and y-axes applies.

Everything is known in these equations except $\mathbf{v}'_2$ and θ_2, which we need to find. We can find two unknowns because we have two independent equations—the equations describing the conservation of momentum in the x and y directions.

Solution

First, we'll solve both conservation of momentum equations ($m_1\mathbf{v}_1 = m_1\mathbf{v}'_1 \cos\theta_1 + m_2\mathbf{v}'_2 \cos\theta_2$ and $0 = m_1\mathbf{v}'_1 \sin\theta_1 + m_2\mathbf{v}'_2 \sin\theta_2$) for $\mathbf{v}'_2 \sin\theta_2$.

For conservation of momentum along x-axis, let's substitute $\sin\theta_2 / \tan\theta_2$ for $\cos\theta_2$ so that terms may cancel out later on. This comes from rearranging the definition of the trigonometric identity $\tan\theta = \sin\theta / \cos\theta$. This gives us

$$m_1\mathbf{v}_1 = m_1\mathbf{v}'_1 \cos\theta_1 + m_2\mathbf{v}'_2 \frac{\sin\theta_2}{\tan\theta_2}. \tag{8.14}$$

Solving for $\mathbf{v}'_2 \sin\theta_2$ yields

$$\mathbf{v}'_2 \sin\theta_2 = \frac{(m_1\mathbf{v}_1 - m_1\mathbf{v}'_1 \cos\theta_1)(\tan\theta_2)}{m_2}. \tag{8.15}$$

For conservation of momentum along y-axis, solving for $\mathbf{v}'_2 \sin\theta_2$ yields

$$\mathbf{v}'_2 \sin\theta_2 = \frac{-(m_1\mathbf{v}'_1 \sin\theta_1)}{m_2}. \tag{8.16}$$

Since both equations equal $\mathbf{v}'_2 \sin\theta_2$, we can set them equal to one another, yielding

$$\frac{(m_1\mathbf{v}_1 - m_1\mathbf{v}'_1\cos\theta_1)(\tan\theta_2)}{m_2} = \frac{-(m_1\mathbf{v}'_1\sin\theta_1)}{m_2}. \tag{8.17}$$

Solving this equation for $\tan\theta_2$, we get

$$\tan\theta_2 = \frac{\mathbf{v}'_1\sin\theta_1}{\mathbf{v}'_1\cos\theta_1 - \mathbf{v}_1}. \tag{8.18}$$

Entering known values into the previous equation gives

$$\tan\theta_2 = \frac{(1.50)(0.707)}{(1.50)(0.707) - 2.00} = -1.129\ . \tag{8.19}$$

Therefore,

$$\theta_2 = \tan^{-1}(-1.129) = 312^0. \tag{8.20}$$

Since angles are defined as positive in the counterclockwise direction, m_2 is scattered to the right.

We'll use the conservation of momentum along the y-axis equation to solve for $\mathbf{v}'_2$.

$$\mathbf{v}'_2 = -\frac{m_1}{m_2}\mathbf{v}'_1\frac{\sin\theta_1}{\sin\theta_2} \tag{8.21}$$

Entering known values into this equation gives

$$\mathbf{v}'_2 = -\frac{(0.250)}{(0.400)}(1.50)\left(\frac{0.7071}{-0.7485}\right). \tag{8.22}$$

Therefore,

$$\mathbf{v}'_2 = 0.886\ \text{m/s}. \tag{8.23}$$

Discussion

Either equation for the *x*- or *y*-axis could have been used to solve for $\mathbf{v}'_2$, but the equation for the *y*-axis is easier because it has fewer terms.

Practice Problems

10. In an elastic collision, an object with momentum $25\ \text{kg}\cdot\text{m/s}$ collides with another object moving to the right that has a momentum $35\ \text{kg}\cdot\text{m/s}$. After the collision, both objects are still moving to the right, but the first object's momentum changes to $10\ \text{kg}\cdot\text{m/s}$. What is the final momentum of the second object?

a. $10\ \text{kg}\cdot\text{m/s}$
b. $20\ \text{kg}\cdot\text{m/s}$
c. $35\ \text{kg}\cdot\text{m/s}$
d. $50\ \text{kg}\cdot\text{m/s}$

11. In an elastic collision, an object with momentum 25 kg · m/s collides with another that has a momentum 35 kg · m/s. The first object's momentum changes to 10 kg · m/s. What is the final momentum of the second object?

a. 10 kg · m/s
b. 20 kg · m/s
c. 35 kg · m/s
d. 50 kg · m/s

Check Your Understanding

12. What is an elastic collision?

a. An elastic collision is one in which the objects after impact are deformed permanently.
b. An elastic collision is one in which the objects after impact lose some of their internal kinetic energy.

c. An elastic collision is one in which the objects after impact do not lose any of their internal kinetic energy.
d. An elastic collision is one in which the objects after impact become stuck together and move with a common velocity.

13. Are perfectly elastic collisions possible?
a. Perfectly elastic collisions are not possible.
b. Perfectly elastic collisions are possible only with subatomic particles.
c. Perfectly elastic collisions are possible only when the objects stick together after impact.
d. Perfectly elastic collisions are possible if the objects and surfaces are nearly frictionless.

14. What is the equation for conservation of momentum for two objects in a one-dimensional collision?
a. $\mathbf{p}_1 + \mathbf{p}_1' = \mathbf{p}_2 + \mathbf{p}_2'$
b. $\mathbf{p}_1 + \mathbf{p}_2 = \mathbf{p}_1' + \mathbf{p}_2'$
c. $\mathbf{p}_1 - \mathbf{p}_2 = \mathbf{p}_1' - \mathbf{p}_2'$
d. $\mathbf{p}_1 + \mathbf{p}_2 + \mathbf{p}_1' + \mathbf{p}_2' = 0$

KEY TERMS

angular momentum the product of the moment of inertia and angular velocity

change in momentum the difference between the final and initial values of momentum; the mass times the change in velocity

elastic collision collision in which objects separate after impact and kinetic energy is conserved

impulse average net external force multiplied by the time the force acts; equal to the change in momentum

impulse–momentum theorem the impulse, or change in momentum, is the product of the net external force and the time over which the force acts

inelastic collision collision in which objects stick together after impact and kinetic energy is not conserved

isolated system system in which the net external force is zero

law of conservation of momentum when the net external force is zero, the total momentum of the system is conserved or constant

linear momentum the product of a system's mass and velocity

point masses structureless particles that cannot rotate or spin

recoil backward movement of an object caused by the transfer of momentum from another object in a collision

SECTION SUMMARY

8.1 Linear Momentum, Force, and Impulse

- Linear momentum, often referenced as *momentum* for short, is defined as the product of a system's mass multiplied by its velocity, $\mathbf{p} = m\mathbf{v}$.
- The SI unit for momentum is kg m/s.
- Newton's second law of motion in terms of momentum states that the net external force equals the change in momentum of a system divided by the time over which it changes, $\mathbf{F}_{\text{net}} = \frac{\Delta \mathbf{p}}{\Delta t}$.
- Impulse is the average net external force multiplied by the time this force acts, and impulse equals the change in momentum, $\Delta \mathbf{p} = \mathbf{F}_{\text{net}} \Delta t$.
- Forces are usually not constant over a period of time, so we use the average of the force over the time it acts.

8.2 Conservation of Momentum

- The law of conservation of momentum is written $\mathbf{p}_{\text{tot}}$ = constant or $\mathbf{p}_{\text{tot}} = \mathbf{p}'_{\text{tot}}$ (isolated system), where $\mathbf{p}_{\text{tot}}$ is the initial total momentum and $\mathbf{p}'_{\text{tot}}$ is the total momentum some time later.
- In an isolated system, the net external force is zero.
- Conservation of momentum applies only when the net external force is zero, within the defined system.

8.3 Elastic and Inelastic Collisions

- If objects separate after impact, the collision is elastic; If they stick together, the collision is inelastic.
- Kinetic energy is conserved in an elastic collision, but not in an inelastic collision.
- The approach to two-dimensional collisions is to choose a convenient coordinate system and break the motion into components along perpendicular axes. Choose a coordinate system with the *x*-axis parallel to the velocity of the incoming particle.
- Two-dimensional collisions of point masses, where mass 2 is initially at rest, conserve momentum along the initial direction of mass 1, or the *x*-axis, and along the direction perpendicular to the initial direction, or the *y*-axis.
- Point masses are structureless particles that cannot spin.

KEY EQUATIONS

8.1 Linear Momentum, Force, and Impulse

impulse	$\mathbf{F}_{\text{net}} \Delta t$
impulse–momentum theorem	$\Delta \mathbf{p} = \mathbf{F}_{\text{net}} \Delta t$
linear momentum	$\mathbf{p} = m\mathbf{v}$
Newton's second law in terms of momentum	$\mathbf{F}_{\text{net}} = \frac{\Delta \mathbf{p}}{\Delta t}$

8.2 Conservation of Momentum

law of conservation of momentum	$\mathbf{p}_{\text{tot}}$ = constant, or $\mathbf{p}_{\text{tot}} = \mathbf{p}'_{\text{tot}}$

conservation of momentum for two objects	$\mathbf{p}_1 + \mathbf{p}_2 = \text{constant}$, or $\mathbf{p}_1 + \mathbf{p}_2 = \mathbf{p}'_1 + \mathbf{p}'_2$
angular momentum	$\mathbf{L} = I\boldsymbol{\omega}$

8.3 Elastic and Inelastic Collisions

conservation of momentum in an elastic collision	$m_1\mathbf{v}_1 + m_2\mathbf{v}_2 = m_1\mathbf{v}'_1 + m_2\mathbf{v}'_2,$
conservation of momentum in an inelastic collision	$m_1\mathbf{v}_1 + m_2\mathbf{v}_2 = (m_1 + m_2)\mathbf{v}'$
conservation of momentum along x-axis for 2D collisions	$m_1\mathbf{v}_1 = m_1\mathbf{v}'_1\cos\theta_1 + m_2\mathbf{v}'_2\cos\theta_2$
conservation of momentum along y-axis for 2D collisions	$0 = m_1\mathbf{v}'_1\sin\theta_1 + m_2\mathbf{v}'_2\sin\theta_2$

CHAPTER REVIEW

Concept Items

8.1 Linear Momentum, Force, and Impulse

1. What is impulse?
   a. Change in velocity
   b. Change in momentum
   c. Rate of change of velocity
   d. Rate of change of momentum

2. In which equation of Newton's second law is mass assumed to be constant?
   a. $\mathbf{F} = ma$
   b. $\mathbf{F} = \frac{\Delta\mathbf{p}}{\Delta t}$
   c. $\mathbf{F} = \Delta\mathbf{p}\Delta t$
   d. $\mathbf{F} = \frac{\Delta m}{\Delta a}$

3. What is the SI unit of momentum?
   a. N
   b. kg · m
   c. kg · m/s
   d. kg · m/s^2

4. What is the equation for linear momentum?
   a. $\mathbf{p} = m\mathbf{v}$
   b. $\mathbf{p} = m/\mathbf{v}$
   c. $\mathbf{p} = \mathrm{m}\mathbf{v}^2$
   d. $\mathbf{p} = \frac{1}{2}\mathrm{m}\mathbf{v}^2$

8.2 Conservation of Momentum

5. What is angular momentum?
   a. The sum of moment of inertia and angular velocity
   b. The ratio of moment of inertia to angular velocity
   c. The product of moment of inertia and angular velocity
   d. Half the product of moment of inertia and square of angular velocity

6. What is an isolated system?
   a. A system in which the net internal force is zero
   b. A system in which the net external force is zero
   c. A system in which the net internal force is a nonzero constant
   d. A system in which the net external force is a nonzero constant

8.3 Elastic and Inelastic Collisions

7. In the equation $\mathbf{p}_1 + \mathbf{p}_2 = \mathbf{p}'_1 + \mathbf{p}'_2$ for the collision of two objects, what is the assumption made regarding the friction acting on the objects?
   a. Friction is zero.
   b. Friction is nearly zero.
   c. Friction acts constantly.
   d. Friction before and after the impact remains the same.

8. What is an inelastic collision?

a. when objects stick together after impact, and their internal energy is not conserved
b. when objects stick together after impact, and their internal energy is conserved
c. when objects stick together after impact, and always come to rest instantaneously after collision
d. when objects stick together after impact, and their internal energy increases

Critical Thinking Items

8.1 Linear Momentum, Force, and Impulse

9. Consider two objects of the same mass. If a force of $100\ \text{N}$ acts on the first for a duration of $1\ \text{s}$ and on the other for a duration of $2\ \text{s}$, which of the following statements is true?
a. The first object will acquire more momentum.
b. The second object will acquire more momentum.
c. Both objects will acquire the same momentum.
d. Neither object will experience a change in momentum.

10. Cars these days have parts that can crumple or collapse in the event of an accident. How does this help protect the passengers?
a. It reduces injury to the passengers by increasing the time of impact.
b. It reduces injury to the passengers by decreasing the time of impact.
c. It reduces injury to the passengers by increasing the change in momentum.
d. It reduces injury to the passengers by decreasing the change in momentum.

11. How much force would be needed to cause a 17 kg · m/s change in the momentum of an object, if the force acted for 5 seconds?
a. 3.4 N
b. 12 N
c. 22 N
d. 85 N

8.2 Conservation of Momentum

12. A billiards ball rolling on the table has momentum $\mathbf{p}_1$. It hits another stationary ball, which then starts rolling. Considering friction to be negligible, what will happen to the momentum of the first ball?
a. It will decrease.
b. It will increase.
c. It will become zero.
d. It will remain the same.

13. A ball rolling on the floor with momentum $\mathbf{p}_1$ collides with a stationary ball and sets it in motion. The momentum of the first ball becomes $\mathbf{p}'_1$, and that of the second becomes $\mathbf{p}'_2$. Compare the magnitudes of $\mathbf{p}_1$ and $\mathbf{p}'_2$.
a. Momenta $\mathbf{p}_1$ and $\mathbf{p}'_2$ are the same in magnitude.
b. The sum of the magnitudes of $\mathbf{p}_1$ and $\mathbf{p}'_2$ is zero.
c. The magnitude of $\mathbf{p}_1$ is greater than that of $\mathbf{p}'_2$.
d. The magnitude of $\mathbf{p}'_2$ is greater than that of $\mathbf{p}_1$.

14. Two cars are moving in the same direction. One car with momentum $\mathbf{p}_1$ collides with another, which has momentum $\mathbf{p}_2$. Their momenta become $\mathbf{p}'_1$ and $\mathbf{p}'_2$ respectively. Considering frictional losses, compare $(\mathbf{p}'_1 + \mathbf{p}'_2)$ with $(\mathbf{p}_1 + \mathbf{p}_2)$.
a. The value of $(\mathbf{p}'_1 + \mathbf{p}'_2)$ is zero.
b. The values of $(\mathbf{p}_1 + \mathbf{p}_2)$ and $(\mathbf{p}'_1 + \mathbf{p}'_2)$ are equal.
c. The value of $(\mathbf{p}_1 + \mathbf{p}_2)$ will be greater than $(\mathbf{p}'_1 + \mathbf{p}'_2)$.
d. The value of $(\mathbf{p}'_1 + \mathbf{p}'_2)$ will be greater than $(\mathbf{p}_1 + \mathbf{p}_2)$.

8.3 Elastic and Inelastic Collisions

15. Two people, who have the same mass, throw two different objects at the same velocity. If the first object is heavier than the second, compare the velocities gained by the two people as a result of recoil.
a. The first person will gain more velocity as a result of recoil.
b. The second person will gain more velocity as a result of recoil.
c. Both people will gain the same velocity as a result of recoil.
d. The velocity of both people will be zero as a result of recoil.

Problems

8.1 Linear Momentum, Force, and Impulse

16. If a force of $50\ \text{N}$ is applied to an object for $0.2\ \text{s}$, and it changes its velocity by $10\ \text{m/s}$, what could be the mass of the object?
a. $1\ \text{kg}$
b. $2\ \text{kg}$
c. $5\ \text{kg}$
d. $250\ \text{kg}$

17. For how long should a force of 130 N be applied to an object of mass 50 kg to change its speed from 20 m/s to 60 m/s?
a. 0.031 s
b. 0.065 s
c. 15.4 s

d. 40 s

8.3 Elastic and Inelastic Collisions

18. If a man with mass 70 kg, standing still, throws an object with mass 5 kg at 50 m/s, what will be the recoil velocity of the man, assuming he is standing on a frictionless surface?
a. –3.6 m/s
b. 0 m/s
c. 3.6 m/s
d. 50.0 m/s

19. Find the recoil velocity of a 65 kg ice hockey goalie who catches a 0.15 kg hockey puck slapped at him at a velocity of 50 m/s. Assume that the goalie is at rest before catching the puck, and friction between the ice and the puck-goalie system is negligible.
a. –0.12 m/s
b. 0 m/s
c. 0.12 m/s
d. 7.5 m/s

Performance Task

8.3 Elastic and Inelastic Collisions

20. You will need the following:
- balls of different weights
- a ruler or wooden strip
- some books
- a paper cup

Make an inclined plane by resting one end of a ruler on a stack of books. Place a paper cup on the other end. Roll a ball from the top of the ruler so that it hits the paper cup. Measure the displacement of the paper cup due to the collision. Now use increasingly heavier balls for this activity and see how that affects the displacement of the cup. Plot a graph of mass vs. displacement. Now repeat the same activity, but this time, instead of using different balls, change the incline of the ruler by varying the height of the stack of books. This will give you different velocities of the ball. See how this affects the displacement of the paper cup.

TEST PREP

Multiple Choice

8.1 Linear Momentum, Force, and Impulse

21. What kind of quantity is momentum?
a. Scalar
b. Vector

22. When does the net force on an object increase?
a. When $\Delta\mathbf{p}$ decreases
b. When Δt increases
c. When Δt decreases

23. In the equation $\Delta\mathbf{p} = m(\mathbf{v}_f - \mathbf{v}_i)$, which quantity is considered to be constant?
a. Initial velocity
b. Final velocity
c. Mass
d. Momentum

24. For how long should a force of 50 N be applied to change the momentum of an object by $12\ \text{kg} \cdot \text{m/s}$?
a. 0.24 s
b. 4.15 s
c. 62 s
d. 600 s

8.2 Conservation of Momentum

25. In the equation $\mathbf{L} = I\boldsymbol{\omega}$, what is I?
a. Linear momentum
b. Angular momentum
c. Torque
d. Moment of inertia

26. Give an example of an isolated system.
a. A cyclist moving along a rough road
b. A figure skater gliding in a straight line on an ice rink
c. A baseball player hitting a home run
d. A man drawing water from a well

8.3 Elastic and Inelastic Collisions

27. In which type of collision is kinetic energy conserved?
a. Elastic
b. Inelastic

28. In physics, what are structureless particles that cannot rotate or spin called?
a. Elastic particles
b. Point masses
c. Rigid masses

29. Two objects having equal masses and velocities collide with each other and come to a rest. What type of a collision is this and why?
a. Elastic collision, because internal kinetic energy is conserved

b. Inelastic collision, because internal kinetic energy is not conserved
c. Elastic collision, because internal kinetic energy is not conserved
d. Inelastic collision, because internal kinetic energy is conserved

30. Two objects having equal masses and velocities collide with each other and come to a rest. Is momentum conserved in this case?
a. Yes
b. No

Short Answer

8.1 Linear Momentum, Force, and Impulse

31. If an object's velocity is constant, what is its momentum proportional to?
a. Its shape
b. Its mass
c. Its length
d. Its breadth

32. If both mass and velocity of an object are constant, what can you tell about its impulse?
a. Its impulse would be constant.
b. Its impulse would be zero.
c. Its impulse would be increasing.
d. Its impulse would be decreasing.

33. When the momentum of an object increases with respect to time, what is true of the net force acting on it?
a. It is zero, because the net force is equal to the rate of change of the momentum.
b. It is zero, because the net force is equal to the product of the momentum and the time interval.
c. It is nonzero, because the net force is equal to the rate of change of the momentum.
d. It is nonzero, because the net force is equal to the product of the momentum and the time interval.

34. How can you express impulse in terms of mass and velocity when neither of those are constant?
a. $\Delta\mathbf{p} = \Delta(m\mathbf{v})$
b. $\frac{\Delta\mathbf{p}}{\Delta t} = \frac{\Delta(m\mathbf{v})}{\Delta t}$
c. $\Delta\mathbf{p} = \Delta(\frac{m}{\mathbf{v}})$
d. $\frac{\Delta\mathbf{p}}{\Delta t} = \frac{1}{\Delta t}\cdot\Delta(m\mathbf{v})$

35. How can you express impulse in terms of mass and initial and final velocities?
a. $\Delta\mathbf{p} = m(\mathbf{v}_\text{f} - \mathbf{v}_\text{i})$
b. $\frac{\Delta\mathbf{p}}{\Delta t} = \frac{m(\mathbf{v}_\text{f}-\mathbf{v}_\text{i})}{\Delta t}$
c. $\Delta\mathbf{p} = \frac{(\mathbf{v}_\text{f}-\mathbf{v}_\text{i})}{m}$
d. $\frac{\Delta\mathbf{p}}{\Delta t} = \frac{1}{m}\frac{(\mathbf{v}_\text{f}-\mathbf{v}_\text{i})}{\Delta t}$

36. Why do we use average force while solving momentum problems? How is net force related to the momentum of the object?
a. Forces are usually constant over a period of time, and net force acting on the object is equal to the rate of change of the momentum.
b. Forces are usually not constant over a period of time, and net force acting on the object is equal to the product of the momentum and the time interval.
c. Forces are usually constant over a period of time, and net force acting on the object is equal to the product of the momentum and the time interval.
d. Forces are usually not constant over a period of time, and net force acting on the object is equal to the rate of change of the momentum.

8.2 Conservation of Momentum

37. Under what condition(s) is the angular momentum of a system conserved?
a. When net torque is zero
b. When net torque is not zero
c. When moment of inertia is constant
d. When both moment of inertia and angular momentum are constant

38. If the moment of inertia of an isolated system increases, what happens to its angular velocity?
a. It increases.
b. It decreases.
c. It stays constant.
d. It becomes zero.

39. If both the moment of inertia and the angular velocity of a system increase, what must be true of the force acting on the system?
a. Force is zero.
b. Force is not zero.
c. Force is constant.
d. Force is decreasing.

8.3 Elastic and Inelastic Collisions

40. Two objects collide with each other and come to a rest. How can you use the equation of conservation of momentum to describe this situation?
a. $m_1\mathbf{v}_1 + m_2\mathbf{v}_2 = 0$
b. $m_1\mathbf{v}_1 - m_2\mathbf{v}_2 = 0$
c. $m_1\mathbf{v}_1 + m_2\mathbf{v}_2 = m_1\mathbf{v}_1'$
d. $m_1\mathbf{v}_1 + m_2\mathbf{v}_2 = m_1\mathbf{v}_2$

41. What is the difference between momentum and impulse?
   a. Momentum is the sum of mass and velocity. Impulse is the change in momentum.
   b. Momentum is the sum of mass and velocity. Impulse is the rate of change in momentum.
   c. Momentum is the product of mass and velocity. Impulse is the change in momentum.
   d. Momentum is the product of mass and velocity. Impulse is the rate of change in momentum.

42. What is the equation for conservation of momentum along the *x*-axis for 2D collisions in terms of mass and velocity, where one of the particles is initially at rest?
   a. $m_1\mathbf{v}_1 = m_1\mathbf{v}_1'\cos\theta_1$
   b. $m_1\mathbf{v}_1 = m_1\mathbf{v}_1'\cos\theta_1 + m_2\mathbf{v}_2'\cos\theta_2$
   c. $m_1\mathbf{v}_1 = m_1\mathbf{v}_1'\cos\theta_1 - m_2\mathbf{v}_2'\cos\theta_2$
   d. $m_1\mathbf{v}_1 = m_1\mathbf{v}_1'\sin\theta_1 + m_2\mathbf{v}_2'\sin\theta_2$

43. What is the equation for conservation of momentum along the *y*-axis for 2D collisions in terms of mass and velocity, where one of the particles is initially at rest?
   a. $0 = m_1\mathbf{v}_1'\sin\theta_1$
   b. $0 = m_1\mathbf{v}_1'\sin\theta_1 + m_2\mathbf{v}_2'\sin\theta_2$
   c. $0 = m_1\mathbf{v}_1'\sin\theta_1 - m_2\mathbf{v}_2'\sin\theta_2$
   d. $0 = m_1\mathbf{v}_1'\cos\theta_1 + m_2\mathbf{v}_2'\cos\theta_2$

Extended Response

8.1 Linear Momentum, Force, and Impulse

44. Can a lighter object have more momentum than a heavier one? How?
   a. No, because momentum is independent of the velocity of the object.
   b. No, because momentum is independent of the mass of the object.
   c. Yes, if the lighter object's velocity is considerably high.
   d. Yes, if the lighter object's velocity is considerably low.

45. Why does it hurt less when you fall on a softer surface?
   a. The softer surface increases the duration of the impact, thereby reducing the effect of the force.
   b. The softer surface decreases the duration of the impact, thereby reducing the effect of the force.
   c. The softer surface increases the duration of the impact, thereby increasing the effect of the force.
   d. The softer surface decreases the duration of the impact, thereby increasing the effect of the force.

46. Can we use the equation $F_{net} = \frac{\Delta p}{\Delta t}$ when the mass is constant?
   a. No, because the given equation is applicable for the variable mass only.
   b. No, because the given equation is not applicable for the constant mass.
   c. Yes, and the resultant equation is $F = m\mathbf{v}$
   d. Yes, and the resultant equation is $F = ma$

8.2 Conservation of Momentum

47. Why does a figure skater spin faster if he pulls his arms and legs in?
   a. Due to an increase in moment of inertia
   b. Due to an increase in angular momentum
   c. Due to conservation of linear momentum
   d. Due to conservation of angular momentum

8.3 Elastic and Inelastic Collisions

48. A driver sees another car approaching him from behind. He fears it is going to collide with his car. Should he speed up or slow down in order to reduce damage?
   a. He should speed up.
   b. He should slow down.
   c. He should speed up and then slow down just before the collision.
   d. He should slow down and then speed up just before the collision.

49. What approach would you use to solve problems involving 2D collisions?
   a. Break the momenta into components and then choose a coordinate system.
   b. Choose a coordinate system and then break the momenta into components.
   c. Find the total momenta in the x and y directions, and then equate them to solve for the unknown.
   d. Find the sum of the momenta in the x and y directions, and then equate it to zero to solve for the unknown.

CHAPTER 9
Work, Energy, and Simple Machines

Figure 9.1 People on a roller coaster experience thrills caused by changes in types of energy. (Jonrev, Wikimedia Commons)

Chapter Outline

9.1 Work, Power, and the Work–Energy Theorem

9.2 Mechanical Energy and Conservation of Energy

9.3 Simple Machines

INTRODUCTION Roller coasters have provided thrills for daring riders around the world since the nineteenth century. Inventors of roller coasters used simple physics to build the earliest examples using railroad tracks on mountainsides and old mines. Modern roller coaster designers use the same basic laws of physics to create the latest amusement park favorites. Physics principles are used to engineer the machines that do the work to lift a roller coaster car up its first big incline before it is set loose to roll. Engineers also have to understand the changes in the car's energy that keep it speeding over hills, through twists, turns, and even loops.

What exactly is energy? How can changes in force, energy, and simple machines move objects like roller coaster cars? How can machines help us do work? In this chapter, you will discover the answer to this question and many more, as you learn about

work, energy, and simple machines.

9.1 Work, Power, and the Work–Energy Theorem

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Describe and apply the work–energy theorem
- Describe and calculate work and power

Section Key Terms

energy	gravitational potential energy	joule	kinetic energy	mechanical energy
potential energy	power	watt	work	work–energy theorem

The Work–Energy Theorem

In physics, the term **work** has a very specific definition. Work is application of force, $\mathbf{f}$, to move an object over a distance, d, in the direction that the force is applied. Work, W, is described by the equation

$$W = \mathbf{f}d.$$

Some things that we typically consider to be work are not work in the scientific sense of the term. Let's consider a few examples. Think about why each of the following statements is true.

- Homework *is not* work.
- Lifting a rock upwards off the ground *is* work.
- Carrying a rock in a straight path across the lawn at a constant speed *is not* work.

The first two examples are fairly simple. Homework is not work because objects are not being moved over a distance. Lifting a rock up off the ground is work because the rock is moving in the direction that force is applied. The last example is less obvious. Recall from the laws of motion that force is *not* required to move an object at constant velocity. Therefore, while some force may be applied to keep the rock up off the ground, no net force is applied to keep the rock moving forward at constant velocity.

Work and **energy** are closely related. When you do work to move an object, you change the object's energy. You (or an object) also expend energy to do work. In fact, energy can be defined as the ability to do work. Energy can take a variety of different forms, and one form of energy can transform to another. In this chapter we will be concerned with **mechanical energy**, which comes in two forms: **kinetic energy** and **potential energy**.

- Kinetic energy is also called energy of motion. A moving object has kinetic energy.
- Potential energy, sometimes called stored energy, comes in several forms. **Gravitational potential energy** is the stored energy an object has as a result of its position above Earth's surface (or another object in space). A roller coaster car at the top of a hill has gravitational potential energy.

Let's examine how doing work on an object changes the object's energy. If we apply force to lift a rock off the ground, we increase the rock's potential energy, PE. If we drop the rock, the force of gravity increases the rock's kinetic energy as the rock moves downward until it hits the ground.

The force we exert to lift the rock is equal to its weight, w, which is equal to its mass, m, multiplied by acceleration due to gravity, $\mathbf{g}$.

$$\mathbf{f} = w = m\mathbf{g}$$

The work we do on the rock equals the force we exert multiplied by the distance, d, that we lift the rock. The work we do on the rock also equals the rock's gain in gravitational potential energy, PE_e.

$$W = PE_e = \mathbf{f}m\mathbf{g}$$

Kinetic energy depends on the mass of an object and its velocity, $\mathbf{v}$.

$$KE = \frac{1}{2}m\mathbf{v}^2$$

When we drop the rock the force of gravity causes the rock to fall, giving the rock kinetic energy. When work done on an object increases only its kinetic energy, then the net work equals the change in the value of the quantity $\frac{1}{2}m\mathbf{v}^2$. This is a statement of the **work–energy theorem**, which is expressed mathematically as

$$W = \Delta KE = \frac{1}{2}m\mathbf{v}_2^2 - \frac{1}{2}m\mathbf{v}_1^2.$$

The subscripts $_2$ and $_1$ indicate the final and initial velocity, respectively. This theorem was proposed and successfully tested by James Joule, shown in Figure 9.2.

Does the name Joule sound familiar? The **joule** (J) is the metric unit of measurement for both work and energy. The measurement of work and energy with the same unit reinforces the idea that work and energy are related and can be converted into one another. 1.0 J = 1.0 N•m, the units of force multiplied by distance. 1.0 N = 1.0 k•m/s^2, so 1.0 J = 1.0 k•m^2/s^2. Analyzing the units of the term (1/2)$m\mathbf{v}^2$ will produce the same units for joules.

Figure 9.2 The joule is named after physicist James Joule (1818–1889). (C. H. Jeens, Wikimedia Commons)

WATCH PHYSICS

Work and Energy

This video explains the work energy theorem and discusses how work done on an object increases the object's KE.

Click to view content (https://www.khanacademy.org/embed_video?v=2WS1sG9fhOk)

GRASP CHECK

True or false—The energy increase of an object acted on only by a gravitational force is equal to the product of the object's weight and the distance the object falls.

a. True
b. False

Calculations Involving Work and Power

In applications that involve work, we are often interested in how fast the work is done. For example, in roller coaster design, the amount of time it takes to lift a roller coaster car to the top of the first hill is an important consideration. Taking a half hour on the ascent will surely irritate riders and decrease ticket sales. Let's take a look at how to calculate the time it takes to do work.

Recall that a rate can be used to describe a quantity, such as work, over a period of time. **Power** is the rate at which work is done. In this case, rate means *per unit of time*. Power is calculated by dividing the work done by the time it took to do the work.

$$P = \frac{W}{t}$$

Let's consider an example that can help illustrate the differences among work, force, and power. Suppose the woman in Figure 9.3 lifting the TV with a pulley gets the TV to the fourth floor in two minutes, and the man carrying the TV up the stairs takes five

minutes to arrive at the same place. They have done the same amount of work ($\mathbf{f}d$) on the TV, because they have moved the same mass over the same vertical distance, which requires the same amount of upward force. However, the woman using the pulley has generated more power. This is because she did the work in a shorter amount of time, so the denominator of the power formula, t, is smaller. (For simplicity's sake, we will leave aside for now the fact that the man climbing the stairs has also done work on himself.)

Figure 9.3 No matter how you move a TV to the fourth floor, the amount of work performed and the potential energy gain are the same.

Power can be expressed in units of **watts** (W). This unit can be used to measure power related to any form of energy or work. You have most likely heard the term used in relation to electrical devices, especially light bulbs. Multiplying power by time gives the amount of energy. Electricity is sold in kilowatt-hours because that equals the amount of electrical energy consumed.

The watt unit was named after James Watt (1736–1819) (see Figure 9.4). He was a Scottish engineer and inventor who discovered how to coax more power out of steam engines.

Figure 9.4 Is James Watt thinking about watts? (Carl Frederik von Breda, Wikimedia Commons)

LINKS TO PHYSICS

Watt's Steam Engine

James Watt did not invent the steam engine, but by the time he was finished tinkering with it, it was more useful. The first steam engines were not only inefficient, they only produced a back and forth, or reciprocal, motion. This was natural because pistons move in and out as the pressure in the chamber changes. This limitation was okay for simple tasks like pumping water or mashing potatoes, but did not work so well for moving a train. Watt was able build a steam engine that converted reciprocal motion to circular motion. With that one innovation, the industrial revolution was off and running. The world would never be the same. One of Watt's steam engines is shown in Figure 9.5. The video that follows the figure explains the importance of the steam engine in the industrial revolution.

Figure 9.5 A late version of the Watt steam engine. (Nehemiah Hawkins, Wikimedia Commons)

WATCH PHYSICS

Watt's Role in the Industrial Revolution

This video demonstrates how the watts that resulted from Watt's inventions helped make the industrial revolution possible and allowed England to enter a new historical era.

Click to view content (https://www.youtube.com/embed/zhL5DCizj5c)

GRASP CHECK

Which form of mechanical energy does the steam engine generate?

a. Potential energy
b. Kinetic energy
c. Nuclear energy
d. Solar energy

Before proceeding, be sure you understand the distinctions among force, work, energy, and power. Force exerted on an object over a distance does work. Work can increase energy, and energy can do work. Power is the rate at which work is done.

WORKED EXAMPLE

Applying the Work–Energy Theorem

An ice skater with a mass of 50 kg is gliding across the ice at a speed of 8 m/s when her friend comes up from behind and gives her a push, causing her speed to increase to 12 m/s. How much work did the friend do on the skater?

Strategy

The work–energy theorem can be applied to the problem. Write the equation for the theorem and simplify it if possible.

$$W = \Delta \text{KE} = \frac{1}{2}m\mathbf{v}_2^2 - \frac{1}{2}m\mathbf{v}_1^2$$

$$\text{Simplify to } W = \frac{1}{2}m(\mathbf{v}_2^2 - \mathbf{v}_1^2)$$

Solution

Identify the variables. $m = 50$ kg,

$$\mathbf{v}_2 = 12\frac{\text{m}}{\text{s}}, \text{ and } \mathbf{v}_1 = 8\frac{\text{m}}{\text{s}} \tag{9.1}$$

Substitute.

$$W = \frac{1}{2}50(12^2 - 8^2) = 2,000 \text{ J} \tag{9.2}$$

Discussion

Work done on an object or system increases its energy. In this case, the increase is to the skater's kinetic energy. It follows that the increase in energy must be the difference in KE before and after the push.

TIPS FOR SUCCESS

This problem illustrates a general technique for approaching problems that require you to apply formulas: Identify the unknown and the known variables, express the unknown variables in terms of the known variables, and then enter all the known values.

Practice Problems

1. How much work is done when a weightlifter lifts a 200 N barbell from the floor to a height of 2 m?
   a. 0 J
   b. 100 J
   c. 200 J
   d. 400 J

2. Identify which of the following actions generates more power. Show your work.

   - carrying a 100 N TV to the second floor in 50 s or
   - carrying a 24 N watermelon to the second floor in 10 s?

   a. Carrying a 100 N TV generates more power than carrying a 24 N watermelon to the same height because power is defined as work done times the time interval.
   b. Carrying a 100 N TV generates more power than carrying a 24 N watermelon to the same height because power is defined as the ratio of work done to the time interval.
   c. Carrying a 24 N watermelon generates more power than carrying a 100 N TV to the same height because power is defined as work done times the time interval.
   d. Carrying a 24 N watermelon generates more power than carrying a 100 N TV to the same height because power is defined as the ratio of work done and the time interval.

Check Your Understanding

3. Identify two properties that are expressed in units of joules.
   a. work and force
   b. energy and weight
   c. work and energy
   d. weight and force

4. When a coconut falls from a tree, work W is done on it as it falls to the beach. This work is described by the equation

$$W = Fd = \frac{1}{2}mv_2^2 - \frac{1}{2}mv_1^2. \quad (9.3)$$

Identify the quantities F, d, m, v_1, and v_2 in this event.

a. F is the force of gravity, which is equal to the weight of the coconut, d is the distance the nut falls, m is the mass of the earth, v_1 is the initial velocity, and v_2 is the velocity with which it hits the beach.
b. F is the force of gravity, which is equal to the weight of the coconut, d is the distance the nut falls, m is the mass of the coconut, v_1 is the initial velocity, and v_2 is the velocity with which it hits the beach.
c. F is the force of gravity, which is equal to the weight of the coconut, d is the distance the nut falls, m is the mass of the earth, v_1 is the velocity with which it hits the beach, and v_2 is the initial velocity.
d. F is the force of gravity, which is equal to the weight of the coconut, d is the distance the nut falls, m is the mass of the coconut, v_1 is the velocity with which it hits the beach, and v_2 is the initial velocity.

9.2 Mechanical Energy and Conservation of Energy

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Explain the law of conservation of energy in terms of kinetic and potential energy
- Perform calculations related to kinetic and potential energy. Apply the law of conservation of energy

Section Key Terms

law of conservation of energy

Mechanical Energy and Conservation of Energy

We saw earlier that mechanical energy can be either potential or kinetic. In this section we will see how energy is transformed from one of these forms to the other. We will also see that, in a closed system, the sum of these forms of energy remains constant.

Quite a bit of potential energy is gained by a roller coaster car and its passengers when they are raised to the top of the first hill. Remember that the *potential* part of the term means that energy has been stored and can be used at another time. You will see that this stored energy can either be used to do work or can be transformed into kinetic energy. For example, when an object that has gravitational potential energy falls, its energy is converted to kinetic energy. Remember that both work and energy are expressed in joules.

Refer back to . The amount of work required to raise the TV from point A to point B is equal to the amount of gravitational potential energy the TV gains from its height above the ground. This is generally true for any object raised above the ground. If all the work done on an object is used to raise the object above the ground, the amount work equals the object's gain in gravitational potential energy. However, note that because of the work done by friction, these energy–work transformations are never perfect. Friction causes the loss of some useful energy. In the discussions to follow, we will use the approximation that transformations are frictionless.

Now, let's look at the roller coaster in Figure 9.6. Work was done on the roller coaster to get it to the top of the first rise; at this point, the roller coaster has gravitational potential energy. It is moving slowly, so it also has a small amount of kinetic energy. As the car descends the first slope, its PE is converted to KE. At the low point much of the original PE has been transformed to KE, and speed is at a maximum. As the car moves up the next slope, some of the KE is transformed back into PE and the car slows down.

Figure 9.6 During this roller coaster ride, there are conversions between potential and kinetic energy.

Virtual Physics

Energy Skate Park Basics

This simulation shows how kinetic and potential energy are related, in a scenario similar to the roller coaster. Observe the changes in *KE* and *PE* by clicking on the bar graph boxes. Also try the three differently shaped skate parks. Drag the skater to the track to start the animation.

Click to view content (http://phet.colorado.edu/sims/html/energy-skate-park-basics/latest/energy-skate-park-basics_en.html)

GRASP CHECK

This simulation (http://phet.colorado.edu/en/simulation/energy-skate-park-basics (http://phet.colorado.edu/en/simulation/energy-skate-park-basics)) shows how kinetic and potential energy are related, in a scenario similar to the roller coaster. Observe the changes in KE and PE by clicking on the bar graph boxes. Also try the three differently shaped skate parks. Drag the skater to the track to start the animation. The bar graphs show how KE and PE are transformed back and forth. Which statement best explains what happens to the mechanical energy of the system as speed is increasing?

a. The mechanical energy of the system increases, provided there is no loss of energy due to friction. The energy would transform to kinetic energy when the speed is increasing.
b. The mechanical energy of the system remains constant provided there is no loss of energy due to friction. The energy would transform to kinetic energy when the speed is increasing.
c. The mechanical energy of the system increases provided there is no loss of energy due to friction. The energy would transform to potential energy when the speed is increasing.
d. The mechanical energy of the system remains constant provided there is no loss of energy due to friction. The energy would transform to potential energy when the speed is increasing.

On an actual roller coaster, there are many ups and downs, and each of these is accompanied by transitions between kinetic and potential energy. Assume that no energy is lost to friction. At any point in the ride, the total mechanical energy is the same, and it is equal to the energy the car had at the top of the first rise. This is a result of the **law of conservation of energy**, which says that, in a closed system, total energy is conserved—that is, it is constant. Using subscripts 1 and 2 to represent initial and final energy, this law is expressed as

$$KE_1 + PE_1 = KE_2 + PE_2.$$

Either side equals the total mechanical energy. The phrase *in a closed system* means we are assuming no energy is lost to the surroundings due to friction and air resistance. If we are making calculations on dense falling objects, this is a good assumption. For the roller coaster, this assumption introduces some inaccuracy to the calculation.

Calculations involving Mechanical Energy and Conservation of Energy

TIPS FOR SUCCESS

When calculating work or energy, use units of meters for distance, newtons for force, kilograms for mass, and seconds for time. This will assure that the result is expressed in joules.

WATCH PHYSICS

Conservation of Energy

This video discusses conversion of *PE* to *KE* and conservation of energy. The scenario is very similar to the roller coaster and the skate park. It is also a good explanation of the energy changes studied in the snap lab.

Click to view content (https://www.khanacademy.org/embed_video?v=kw_4Loo1HR4)

GRASP CHECK

Did you expect the speed at the bottom of the slope to be the same as when the object fell straight down? Which statement best explains why this is not exactly the case in real-life situations?

a. The speed was the same in the scenario in the animation because the object was sliding on the ice, where there is large amount of friction. In real life, much of the mechanical energy is lost as heat caused by friction.
b. The speed was the same in the scenario in the animation because the object was sliding on the ice, where there is small amount of friction. In real life, much of the mechanical energy is lost as heat caused by friction.
c. The speed was the same in the scenario in the animation because the object was sliding on the ice, where there is large amount of friction. In real life, no mechanical energy is lost due to conservation of the mechanical energy.
d. The speed was the same in the scenario in the animation because the object was sliding on the ice, where there is small amount of friction. In real life, no mechanical energy is lost due to conservation of the mechanical energy.

WORKED EXAMPLE

Applying the Law of Conservation of Energy

A 10 kg rock falls from a 20 m cliff. What is the kinetic and potential energy when the rock has fallen 10 m?

Strategy

Choose the equation.

$$KE_1 + PE_1 = KE_2 + PE_2 \tag{9.4}$$

$$KE = \tfrac{1}{2}m\mathbf{v}^2; \quad PE = m\mathbf{g}h \tag{9.5}$$

$$\frac{1}{2}m\mathbf{v}_1^2 + m\mathbf{g}h_1 = \frac{1}{2}m\mathbf{v}_2^2 + m\mathbf{g}h_2 \tag{9.6}$$

List the knowns.

$m = 10$ kg, $\mathbf{v}_1 = 0$, $\mathbf{g} = 9.80$

$$\frac{\mathrm{m}}{\mathrm{s}^2}, \tag{9.7}$$

$h_1 = 20$ m, $h_2 = 10$ m

Identify the unknowns.

KE_2 and PE_2

Substitute the known values into the equation and solve for the unknown variables.

Solution

$$PE_2 = m\mathbf{g}h_2 = 10\,(9.80)\,10 = 980 \text{ J} \tag{9.8}$$

$$KE_2 = PE_2 - (KE_1 + PE_1) = 980 - \{[0 - [10\,(9.80)\,20]]\} = 980 \text{ J} \tag{9.9}$$

Discussion

Alternatively, conservation of energy equation could be solved for $\mathbf{v}_2$ and KE_2 could be calculated. Note that m could also be eliminated.

TIPS FOR SUCCESS

Note that we can solve many problems involving conversion between *KE* and *PE* without knowing the mass of the object in question. This is because kinetic and potential energy are both proportional to the mass of the object. In a situation where $KE = PE$, we know that $m\mathbf{g}h = (1/2)m\mathbf{v}^2$.

Dividing both sides by m and rearranging, we have the relationship

$2\mathbf{g}h = \mathbf{v}^2$.

Practice Problems

5. A child slides down a playground slide. If the slide is 3 m high and the child weighs 300 N, how much potential energy does the child have at the top of the slide? (Round g to $10 \ m/s^2$.)
   a. 0 J
   b. 100 J
   c. 300 J
   d. 900 J

6. A 0.2 kg apple on an apple tree has a potential energy of 10 J. It falls to the ground, converting all of its PE to kinetic energy. What is the velocity of the apple just before it hits the ground?
   a. 0 m/s
   b. 2 m/s
   c. 10 m/s
   d. 50 m/s

Snap Lab

Converting Potential Energy to Kinetic Energy

In this activity, you will calculate the potential energy of an object and predict the object's speed when all that potential energy has been converted to kinetic energy. You will then check your prediction.

You will be dropping objects from a height. Be sure to stay a safe distance from the edge. Don't lean over the railing too far. Make sure that you do not drop objects into an area where people or vehicles pass by. Make sure that dropping objects will not cause damage.

You will need the following:

Materials for each pair of students:
- Four marbles (or similar small, dense objects)
- Stopwatch

Materials for class:
- Metric measuring tape long enough to measure the chosen height
- A scale

Instructions

Procedure

1. Work with a partner. Find and record the mass of four small, dense objects per group.
2. Choose a location where the objects can be safely dropped from a height of at least 15 meters. A bridge over water with a safe pedestrian walkway will work well.
3. Measure the distance the object will fall.
4. Calculate the potential energy of the object before you drop it using $PE = m\mathbf{g}h = (9.80)mh$.
5. Predict the kinetic energy and velocity of the object when it lands using $PE = KE$ and so, $m\mathbf{g}h = \frac{m\mathbf{v}^2}{2}$; $\mathbf{v} = \sqrt{2(9.80)h} = 4.43\sqrt{h}$.
6. One partner drops the object while the other measures the time it takes to fall.
7. Take turns being the dropper and the timer until you have made four measurements.
8. Average your drop multiplied by and calculate the velocity of the object when it landed using $\mathbf{v} = \mathbf{a}t = \mathbf{g}t = (9.80)t$.
9. Compare your results to your prediction.

GRASP CHECK

Galileo's experiments proved that, contrary to popular belief, heavy objects do not fall faster than light objects. How do the equations you used support this fact?

a. Heavy objects do not fall faster than the light objects because while conserving the mechanical energy of the system, the mass term gets cancelled and the velocity is independent of the mass. In real life, the variation in the velocity of the different objects is observed because of the non-zero air resistance.
b. Heavy objects do not fall faster than the light objects because while conserving the mechanical energy of the system, the mass term does not get cancelled and the velocity is dependent on the mass. In real life, the variation in the velocity of the different objects is observed because of the non-zero air resistance.
c. Heavy objects do not fall faster than the light objects because while conserving the mechanical energy the system, the mass term gets cancelled and the velocity is independent of the mass. In real life, the variation in the velocity of the different objects is observed because of zero air resistance.
d. Heavy objects do not fall faster than the light objects because while conserving the mechanical energy of the system, the mass term does not get cancelled and the velocity is dependent on the mass. In real life, the variation in the velocity of the different objects is observed because of zero air resistance.

Check Your Understanding

7. Describe the transformation between forms of mechanical energy that is happening to a falling skydiver before his parachute opens.
   a. Kinetic energy is being transformed into potential energy.
   b. Potential energy is being transformed into kinetic energy.
   c. Work is being transformed into kinetic energy.
   d. Kinetic energy is being transformed into work.

8. True or false—If a rock is thrown into the air, the increase in the height would increase the rock's kinetic energy, and then the increase in the velocity as it falls to the ground would increase its potential energy.
   a. True
   b. False

9. Identify equivalent terms for *stored energy* and *energy of motion*.
   a. Stored energy is potential energy, and energy of motion is kinetic energy.
   b. Energy of motion is potential energy, and stored energy is kinetic energy.
   c. Stored energy is the potential as well as the kinetic energy of the system.
   d. Energy of motion is the potential as well as the kinetic energy of the system.

9.3 Simple Machines

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Describe simple and complex machines
- Calculate mechanical advantage and efficiency of simple and complex machines

Section Key Terms

complex machine	efficiency output	ideal mechanical advantage	inclined plane	input work
lever	mechanical advantage	output work	pulley	screw
simple machine	wedge	wheel and axle		

Simple Machines

Simple machines make work easier, but they do not decrease the amount of work you have to do. Why can't simple machines change the amount of work that you do? Recall that in closed systems the total amount of energy is conserved. A machine cannot increase the amount of energy you put into it. So, why is a simple machine useful? Although it cannot change the amount of work you do, a simple machine can change the amount of force you must apply to an object, and the distance over which you apply the force. In most cases, a simple machine is used to reduce the amount of force you must exert to do work. The down side is that you must exert the force over a greater distance, because the product of force and distance, $\mathbf{f}d$, (which equals work) does not change.

Let's examine how this works in practice. In Figure 9.7(a), the worker uses a type of **lever** to exert a small force over a large distance, while the pry bar pulls up on the nail with a large force over a small distance. Figure 9.7(b) shows the how a lever works mathematically. The effort force, applied at $\mathbf{F_e}$, lifts the load (the resistance force) which is pushing down at $\mathbf{F}_r$. The triangular pivot is called the fulcrum; the part of the lever between the fulcrum and $\mathbf{F}_e$ is the effort arm, L_e; and the part to the left is the resistance arm, L_r. The **mechanical advantage** is a number that tells us how many times a simple machine multiplies the effort force. The **ideal mechanical advantage**, *IMA*, is the mechanical advantage of a perfect machine with no loss of useful work caused by friction between moving parts. The equation for *IMA* is shown in Figure 9.7(b).

Figure 9.7 (a) A pry bar is a type of lever. (b) The ideal mechanical advantage equals the length of the effort arm divided by the length of the resistance arm of a lever.

In general, the *IMA* = the resistance force, $\mathbf{F}_r$, divided by the effort force, $\mathbf{F}_e$. *IMA* also equals the distance over which the effort is applied, d_e, divided by the distance the load travels, d_r.

$$IMA = \frac{\mathbf{F}_r}{\mathbf{F}_e} = \frac{d_e}{d_r}$$

Getting back to conservation of energy, for any simple machine, the work put into the machine, W_i, equals the work the machine puts out, W_o. Combining this with the information in the paragraphs above, we can write

$$W_i = W_o$$
$$\mathbf{F}_e d_e = \mathbf{F}_r d_r$$
$$\text{If } \mathbf{F}_e < \mathbf{F}_r, \text{ then } d_e > d_r.$$

The equations show how a simple machine can output the same amount of work while reducing the amount of effort force by increasing the distance over which the effort force is applied.

WATCH PHYSICS

Introduction to Mechanical Advantage

This video shows how to calculate the *IMA* of a lever by three different methods: (1) from effort force and resistance force; (2) from the lengths of the lever arms, and; (3) from the distance over which the force is applied and the distance the load moves.

Click to view content (https://www.youtube.com/embed/pfzJ-z5Ij48)

GRASP CHECK

Two children of different weights are riding a seesaw. How do they position themselves with respect to the pivot point (the fulcrum) so that they are balanced?

a. The heavier child sits closer to the fulcrum.
b. The heavier child sits farther from the fulcrum.
c. Both children sit at equal distance from the fulcrum.
d. Since both have different weights, they will never be in balance.

Some levers exert a large force to a short effort arm. This results in a smaller force acting over a greater distance at the end of the resistance arm. Examples of this type of lever are baseball bats, hammers, and golf clubs. In another type of lever, the fulcrum is at the end of the lever and the load is in the middle, as in the design of a wheelbarrow.

The simple machine shown in Figure 9.8 is called a **wheel and axle.** It is actually a form of lever. The difference is that the effort arm can rotate in a complete circle around the fulcrum, which is the center of the axle. Force applied to the outside of the wheel causes a greater force to be applied to the rope that is wrapped around the axle. As shown in the figure, the ideal mechanical advantage is calculated by dividing the radius of the wheel by the radius of the axle. Any crank-operated device is an example of a wheel and axle.

Figure 9.8 Force applied to a wheel exerts a force on its axle.

An **inclined plane** and a **wedge** are two forms of the same simple machine. A wedge is simply two inclined planes back to back. Figure 9.9 shows the simple formulas for calculating the *IMA*s of these machines. All sloping, paved surfaces for walking or driving are inclined planes. Knives and axe heads are examples of wedges.

Figure 9.9 An inclined plane is shown on the left, and a wedge is shown on the right.

The **screw** shown in Figure 9.10 is actually a lever attached to a circular inclined plane. Wood screws (of course) are also examples of screws. The lever part of these screws is a screw driver. In the formula for *IMA*, the distance between screw threads is called *pitch* and has the symbol *P*.

Figure 9.10 The screw shown here is used to lift very heavy objects, like the corner of a car or a house a short distance.

Figure 9.11 shows three different **pulley** systems. Of all simple machines, mechanical advantage is easiest to calculate for pulleys. Simply count the number of ropes supporting the load. That is the *IMA*. Once again we have to exert force over a longer distance to multiply force. To raise a load 1 meter with a pulley system you have to pull *N* meters of rope. Pulley systems are often used to raise flags and window blinds and are part of the mechanism of construction cranes.

Figure 9.11 Three pulley systems are shown here.

WATCH PHYSICS

Mechanical Advantage of Inclined Planes and Pulleys

The first part of this video shows how to calculate the *IMA* of pulley systems. The last part shows how to calculate the *IMA* of an inclined plane.

Click to view content (https://www.khanacademy.org/embed_video?v=vSsK7Rfa3yA)

GRASP CHECK

How could you use a pulley system to lift a light load to great height?

a. Reduce the radius of the pulley.
b. Increase the number of pulleys.
c. Decrease the number of ropes supporting the load.
d. Increase the number of ropes supporting the load.

A **complex machine** is a combination of two or more simple machines. The wire cutters in Figure 9.12 combine two levers and two wedges. Bicycles include wheel and axles, levers, screws, and pulleys. Cars and other vehicles are combinations of many machines.

Figure 9.12 Wire cutters are a common complex machine.

Calculating Mechanical Advantage and Efficiency of Simple Machines

In general, the *IMA* = the resistance force, $\mathbf{F}_r$, divided by the effort force, $\mathbf{F}_e$. *IMA* also equals the distance over which the effort is applied, d_e, divided by the distance the load travels, d_r.

$$IMA = \frac{\mathbf{F}_r}{\mathbf{F}_e} = \frac{d_e}{d_r}$$

Refer back to the discussions of each simple machine for the specific equations for the *IMA* for each type of machine.

No simple or complex machines have the actual mechanical advantages calculated by the *IMA* equations. In real life, some of the applied work always ends up as wasted heat due to friction between moving parts. Both the **input work** (W_i) and **output work** (W_o) are the result of a force, $\mathbf{F}$, acting over a distance, d.

$$W_i = \mathbf{F}_i d_i \text{and} W_o = \mathbf{F}_o d_o$$

The **efficiency output** of a machine is simply the output work divided by the input work, and is usually multiplied by 100 so that it is expressed as a percent.

$$\%\ \text{efficiency} = \frac{W_o}{W_i} \times 100$$

Look back at the pictures of the simple machines and think about which would have the highest efficiency. Efficiency is related to friction, and friction depends on the smoothness of surfaces and on the area of the surfaces in contact. How would lubrication affect the efficiency of a simple machine?

WORKED EXAMPLE

Efficiency of a Lever

The input force of 11 N acting on the effort arm of a lever moves 0.4 m, which lifts a 40 N weight resting on the resistance arm a

distance of 0.1 m. What is the efficiency of the machine?

Strategy

State the equation for efficiency of a simple machine, $\%\ \text{efficiency} = \frac{W_o}{W_i} \times 100$, and calculate W_o and W_i. Both work values are the product *Fd*.

Solution

$W_i = \mathbf{F}_i d_i$ = (11)(0.4) = 4.4 J and $W_o = \mathbf{F}_o d_o$ = (40)(0.1) = 4.0 J, then $\%\ \text{efficiency} = \frac{W_o}{W_i} \times 100 = \frac{4.0}{4.4} \times 100 = 91\%$

Discussion

Efficiency in real machines will always be less than 100 percent because of work that is converted to unavailable heat by friction and air resistance. W_o and W_i can always be calculated as a force multiplied by a distance, although these quantities are not always as obvious as they are in the case of a lever.

Practice Problems

10. What is the IMA of an inclined plane that is $5\ \text{m}$ long and $2\ \text{m}$ high?

a. 0.4
b. 2.5
c. $0.4\ \text{m}$
d. $2.5\ \text{m}$

11. If a pulley system can lift a 200N load with an effort force of 52 N and has an efficiency of almost 100 percent, how many ropes are supporting the load?

a. 1 rope is required because the actual mechanical advantage is 0.26.
b. 1 rope is required because the actual mechanical advantage is 3.80.
c. 4 ropes are required because the actual mechanical advantage is 0.26.
d. 4 ropes are required because the actual mechanical advantage is 3.80.

Check Your Understanding

12. True or false—The efficiency of a simple machine is always less than 100 percent because some small fraction of the input work is always converted to heat energy due to friction.

a. True
b. False

13. The circular handle of a faucet is attached to a rod that opens and closes a valve when the handle is turned. If the rod has a diameter of 1 cm and the IMA of the machine is 6, what is the radius of the handle?

A. 0.08 cm
B. 0.17 cm
C. 3.0 cm
D. 6.0 cm

KEY TERMS

complex machine a machine that combines two or more simple machines
efficiency output work divided by input work
energy the ability to do work
gravitational potential energy energy acquired by doing work against gravity
ideal mechanical advantage the mechanical advantage of an idealized machine that loses no energy to friction
inclined plane a simple machine consisting of a slope
input work effort force multiplied by the distance over which it is applied
joule the metric unit for work and energy; equal to 1 newton meter (N•m)
kinetic energy energy of motion
law of conservation of energy states that energy is neither created nor destroyed
lever a simple machine consisting of a rigid arm that pivots on a fulcrum
mechanical advantage the number of times the input force is multiplied
mechanical energy kinetic or potential energy
output work output force multiplied by the distance over which it acts
potential energy stored energy
power the rate at which work is done
pulley a simple machine consisting of a rope that passes over one or more grooved wheels
screw a simple machine consisting of a spiral inclined plane
simple machine a machine that makes work easier by changing the amount or direction of force required to move an object
watt the metric unit of power; equivalent to joules per second
wedge a simple machine consisting of two back-to-back inclined planes
wheel and axle a simple machine consisting of a rod fixed to the center of a wheel
work force multiplied by distance
work–energy theorem states that the net work done on a system equals the change in kinetic energy

SECTION SUMMARY

9.1 Work, Power, and the Work–Energy Theorem

- Doing work on a system or object changes its energy.
- The work–energy theorem states that an amount of work that changes the velocity of an object is equal to the change in kinetic energy of that object.The work–energy theorem states that an amount of work that changes the velocity of an object is equal to the change in kinetic energy of that object.
- Power is the rate at which work is done.

9.2 Mechanical Energy and Conservation of Energy

- Mechanical energy may be either kinetic (energy of motion) or potential (stored energy).
- Doing work on an object or system changes its energy.
- Total energy in a closed, isolated system is constant.

9.3 Simple Machines

- The six types of simple machines make work easier by changing the $\mathbf{f}d$ term so that force is reduced at the expense of increased distance.
- The ratio of output force to input force is a machine's mechanical advantage.
- Combinations of two or more simple machines are called complex machines.
- The ratio of output work to input work is a machine's efficiency.

KEY EQUATIONS

9.1 Work, Power, and the Work–Energy Theorem

equation for work	$W = \mathbf{f}d$
force	$\mathbf{f} = w = m\mathbf{g}$
work equivalencies	$W = PE_e = \mathbf{f}m\mathbf{g}$
kinetic energy	$KE = \frac{1}{2}m\mathbf{v}^2$
work–energy theorem	$W = \Delta\text{KE} = \frac{1}{2}m\mathbf{v}_2^2 - \frac{1}{2}m\mathbf{v}_1^2$
power	$P = \frac{W}{t}$

9.2 Mechanical Energy and Conservation of Energy

conservation of energy	$KE_1 + PE_1 = KE_2 + PE_2$

9.3 Simple Machines

ideal mechanical advantage (general)	$IMA = \frac{\mathbf{F}_r}{\mathbf{F}_e} = \frac{d_e}{d_r}$
ideal mechanical advantage (lever)	$IMA = \frac{L_e}{L_r}$
ideal mechanical advantage (wheel and axle)	$IMA = \frac{R}{r}$
ideal mechanical advantage (inclined plane)	$IMA = \frac{L}{h}$
ideal mechanical advantage (wedge)	$IMA = \frac{L}{t}$
ideal mechanical advantage (pulley)	$IMA = N$
ideal mechanical advantage (screw)	$IMA = \frac{2\pi L}{P}$
input work	$W_i = \mathbf{F}_i d_i$
output work	$W_o = \mathbf{F}_o d_o$
efficiency output	$\%\ \text{efficiency} = \frac{W_o}{W_i} \times 100$

CHAPTER REVIEW

Concept Items

9.1 Work, Power, and the Work–Energy Theorem

1. Is it possible for the sum of kinetic energy and potential energy of an object to change without work having been done on the object? Explain.
   a. No, because the work-energy theorem states that work done on an object is equal to the change in kinetic energy, and change in KE requires a change in velocity. It is assumed that mass is constant.
   b. No, because the work-energy theorem states that work done on an object is equal to the sum of kinetic energy, and the change in KE requires a change in displacement. It is assumed that mass is constant.
   c. Yes, because the work-energy theorem states that work done on an object is equal to the change in kinetic energy, and change in KE requires a change in velocity. It is assumed that mass is constant.
   d. Yes, because the work-energy theorem states that work done on an object is equal to the sum of kinetic energy, and the change in KE requires a change in displacement. It is assumed that mass is constant.

2. Define work for one-dimensional motion.
   a. Work is defined as the ratio of the force over the distance.
   b. Work is defined as the sum of the force and the distance.
   c. Work is defined as the square of the force over the distance.
   d. Work is defined as the product of the force and the distance.

3. A book with a mass of 0.30 kg falls 2 m from a shelf to the floor. This event is described by the work–energy theorem: $W = fd = \frac{1}{2}mv_2^2 - \frac{1}{2}mv_1^2$ Explain why this is enough information to calculate the speed with which the book hits the floor.
   a. The mass of the book, m, and distance, d, are stated. F is the weight of the book mg. v_1 is the initial velocity and v_2 is the final velocity. The final velocity is the only unknown quantity.
   b. The mass of the book, m, and distance, d, are stated. F is the weight of the book mg. v_1 is the final velocity and v_2 is the initial velocity. The final velocity is the only unknown quantity.
   c. The mass of the book, m, and distance, d, are stated. F is the weight of the book mg. v_1 is the initial velocity and v_2 is the final velocity. The final velocity and the initial velocities are the only unknown quantities.
   d. The mass of the book, m, and distance, d, are stated. F is the weight of the book mg. v_1 is the final velocity and v_2 is the initial velocity. The final velocity and the initial velocities are the only unknown quantities.

9.2 Mechanical Energy and Conservation of Energy

4. Describe the changes in KE and PE of a person jumping up and down on a trampoline.
   a. While going up, the person's KE would change to PE. While coming down, the person's PE would change to KE.
   b. While going up, the person's PE would change to KE. While coming down, the person's KE would change to PE.
   c. While going up, the person's KE would not change, but while coming down, the person's PE would change to KE.
   d. While going up, the person's PE would change to KE, but while coming down, the person's KE would not change.

5. You know the height from which an object is dropped. Which equation could you use to calculate the velocity as the object hits the ground?
   a. $v = h$
   b. $v = \sqrt{2h}$
   c. $v = gh$
   d. $v = \sqrt{2gh}$

6. The starting line of a cross country foot race is at the bottom of a hill. Which form(s) of mechanical energy of the runners will change when the starting gun is fired?
   a. Kinetic energy only
   b. Potential energy only
   c. Both kinetic and potential energy
   d. Neither kinetic nor potential energy

9.3 Simple Machines

7. How does a simple machine make work easier?
   a. It reduces the input force and the output force.
   b. It reduces the input force and increases the output force.
   c. It increases the input force and reduces the output force.
   d. It increases the input force and the output force.

8. Which type of simple machine is a knife?
   a. A ramp
   b. A wedge
   c. A pulley
   d. A screw

Critical Thinking Items

9.1 Work, Power, and the Work–Energy Theorem

9. Which activity requires a person to exert force on an object that causes the object to move but does not change the kinetic or potential energy of the object?
   a. Moving an object to a greater height with acceleration
   b. Moving an object to a greater height without acceleration
   c. Carrying an object with acceleration at the same height
   d. Carrying an object without acceleration at the same height

10. Which statement explains how it is possible to carry books to school without changing the kinetic or potential energy of the books or doing any work?
   a. By moving the book without acceleration and keeping the height of the book constant
   b. By moving the book with acceleration and keeping the height of the book constant
   c. By moving the book without acceleration and changing the height of the book
   d. By moving the book with acceleration and changing the height of the book

9.2 Mechanical Energy and Conservation of Energy

11. True or false—A cyclist coasts down one hill and up another hill until she comes to a stop. The point at which the bicycle stops is lower than the point at which it started coasting because part of the original potential energy has been converted to a quantity of heat and this makes the tires of the bicycle warm.
   a. True
   b. False

9.3 Simple Machines

12. We think of levers being used to decrease effort force. Which of the following describes a lever that requires a large effort force which causes a smaller force to act over a large distance and explains how it works?
   a. Anything that is swung by a handle, such as a hammer or racket. Force is applied near the fulcrum over a short distance, which makes the other end move rapidly over a long distance.
   b. Anything that is swung by a handle, such as a hammer or racket. Force is applied far from the fulcrum over a large distance, which makes the other end move rapidly over a long distance.
   c. A lever used to lift a heavy stone. Force is applied near the fulcrum over a short distance, which

makes the other end lift a heavy object easily.

d. A lever used to lift a heavy stone. Force is applied far from the fulcrum over a large distance, which makes the other end lift a heavy object easily

13. A baseball bat is a lever. Which of the following explains how a baseball bat differs from a lever like a pry bar?

a. In a baseball bat, effort force is smaller and is applied over a large distance, while the resistance force is smaller and is applied over a long distance.

b. In a baseball bat, effort force is smaller and is applied over a large distance, while the resistance force is smaller and is applied over a short distance.

c. In a baseball bat, effort force is larger and is applied over a short distance, while the resistance force is smaller and is applied over a long distance.

d. In a baseball bat, effort force is larger and is applied over a short distance, while the resistance force is smaller and is applied over a short distance.

Problems

9.1 Work, Power, and the Work–Energy Theorem

14. A baseball player exerts a force of 100 N on a ball for a distance of 0.5 m as he throws it. If the ball has a mass of 0.15 kg, what is its velocity as it leaves his hand?

a. -36.5 m/s
b. -25.8 m/s
c. 25.8 m/s
d. 36.5 m/s

15. A boy pushes his little sister on a sled. The sled accelerates from 0 to 3.2 m/s . If the combined mass of his sister and the sled is 40.0 kg and 18 W of power were generated, how long did the boy push the sled?

a. 205 s
b. 128 s
c. 23 s
d. 11 s

9.2 Mechanical Energy and Conservation of Energy

16. What is the kinetic energy of a 0.01 kg bullet traveling at a velocity of 700 m/s?

a. 3.5 J
b. 7 J
c. $2.45 \times 10^3\text{ J}$
d. $2.45 \times 10^5\text{ J}$

17. A marble rolling across a flat, hard surface at 2 m/s rolls up a ramp. Assuming that $g = 10\text{ m/s}^2$ and no energy is lost to friction, what will be the vertical height of the marble when it comes to a stop before rolling back down? Ignore effects due to the rotational kinetic energy.

a. 0.1 m
b. 0.2 m
c. 0.4 m
d. 2 m

18. The potential energy stored in a compressed spring is $U = \frac{1}{2}kx^2$, where k is the force constant and x is the distance the spring is compressed from the equilibrium position. Four experimental setups described below can be used to determine the force constant of a spring. Which one(s) require measurement of the fewest number of variables to determine k? Assume the acceleration due to gravity is known.

I. An object is propelled vertically by a compressed spring.
II. An object is propelled horizontally on a frictionless surface by a compressed spring.
III. An object is statically suspended from a spring.
IV. An object suspended from a spring is set into oscillatory motion.

a. I only
b. III only
c. I and II only
d. III and IV only

9.3 Simple Machines

19. A man is using a wedge to split a block of wood by hitting the wedge with a hammer. This drives the wedge into the wood creating a crack in the wood. When he hits the wedge with a force of 400 N it travels 4 cm into the wood. This caused the wedge to exert a force of 1,400 N sideways increasing the width of the crack by 1 cm . What is the efficiency of the wedge?

a. 0.875 percent
b. 0.14
c. 0.751
d. 87.5 percent

20. An electrician grips the handles of a wire cutter, like the one shown, 10 cm from the pivot and places a wire between the jaws 2 cm from the pivot. If the cutter blades are 2 cm wide and 0.3 cm thick, what is the overall IMA of this complex machine?

a. 1.34
b. 1.53
c. 33.3
d. 33.5

Performance Task

9.3 Simple Machines

21. Conservation of Energy and Energy Transfer; Cause and Effect; and S&EP, Planning and Carrying Out Investigations

Plan an investigation to measure the mechanical advantage of simple machines and compare to the *IMA* of the machine. Also measure the efficiency of each machine studied. Design an investigation to make these measurements for these simple machines: lever, inclined plane, wheel and axle and a pulley system. In addition to these machines, include a spring scale, a tape measure, and a weight with a loop on top that can be attached to the hook on the spring scale. A spring scale is shown in the image.

A spring scale measures weight, not mass.

LEVER: Beginning with the lever, explain how you would measure input force, output force, effort arm, and resistance arm. Also explain how you would find the distance the load travels and the distance over which the effort force is applied. Explain how you would use this data to determine *IMA* and efficiency.

INCLINED PLANE: Make measurements to determine *IMA* and efficiency of an inclined plane. Explain how you would use the data to calculate these values. Which property do you already know? Note that there are no effort and resistance arm measurements, but there are height and length measurements.

WHEEL AND AXLE: Again, you will need two force measurements and four distance measurements. Explain how you would use these to calculate *IMA* and efficiency.

SCREW: You will need two force measurements, two distance traveled measurements, and two length measurements. You may describe a screw like the one shown in Figure 9.10 or you could use a screw and screw driver. (Measurements would be easier for the former). Explain how you would use these to calculate *IMA* and efficiency.

PULLEY SYSTEM: Explain how you would determine the *IMA* and efficiency of the four-pulley system shown in Figure 9.11. Why do you only need two distance measurements for this machine?

Design a table that compares the efficiency of the five simple machines. Make predictions as to the most and least efficient machines.

TEST PREP

Multiple Choice

9.1 Work, Power, and the Work–Energy Theorem

22. Which expression represents power?

a. fd
b. mgh
c. $\frac{mv^2}{2}$
d. $\frac{W}{t}$

23. The work–energy theorem states that the change in the kinetic energy of an object is equal to what?

a. The work done on the object
b. The force applied to the object
c. The loss of the object's potential energy
d. The object's total mechanical energy minus its kinetic energy

24. A runner at the start of a race generates 250 W of power as he accelerates to 5 m/s . If the runner has a mass of 60 kg, how long did it take him to reach that speed?

a. 0.33 s
b. 0.83 s
c. 1.2 s
d. 3.0 s

25. A car's engine generates 100,000 W of power as it exerts a force of 10,000 N. How long does it take the car to travel 100 m?
   a. 0.001 s
   b. 0.01 s
   c. 10 s
   d. 1,000 s

9.2 Mechanical Energy and Conservation of Energy

26. Why is this expression for kinetic energy incorrect? $\text{KE} = (m)(v)^2$.
   a. The constant g is missing.
   b. The term v should not be squared.
   c. The expression should be divided by 2.
   d. The energy lost to friction has not been subtracted.

27. What is the kinetic energy of a 10kg object moving at $2.0\ \text{m/s}$?
   a. $10\ \text{J}$
   b. $20\ \text{J}$
   c. $40\ \text{J}$
   d. $100\ \text{J}$

28. Which statement best describes the PE-KE transformations for a javelin, starting from the instant the javelin leaves the thrower's hand until it hits the ground.
   a. Initial PE is transformed to KE until the javelin reaches the high point of its arc. On the way back down, KE is transformed into PE. At every point in the flight, mechanical energy is being transformed into heat energy.
   b. Initial KE is transformed to PE until the javelin reaches the high point of its arc. On the way back down, PE is transformed into KE. At every point in the flight, mechanical energy is being transformed into heat energy.
   c. Initial PE is transformed to KE until the javelin reaches the high point of its arc. On the way back down, there is no transformation of mechanical energy. At every point in the flight, mechanical energy is being transformed into heat energy.
   d. Initial KE is transformed to PE until the javelin reaches the high point of its arc. On the way back down, there is no transformation of mechanical energy. At every point in the flight, mechanical energy is being transformed into heat energy.

29. At the beginning of a roller coaster ride, the roller coaster car has an initial energy mostly in the form of PE. Which statement explains why the fastest speeds of the car will be at the lowest points in the ride?
   a. At the bottom of the slope kinetic energy is at its maximum value and potential energy is at its minimum value.
   b. At the bottom of the slope potential energy is at its maximum value and kinetic energy is at its minimum value.
   c. At the bottom of the slope both kinetic and potential energy reach their maximum values
   d. At the bottom of the slope both kinetic and potential energy reach their minimum values.

9.3 Simple Machines

30. A large radius divided by a small radius is the expression used to calculate the IMA of what?
   a. A screw
   b. A pulley
   c. A wheel and axle
   d. An inclined plane.

31. What is the IMA of a wedge that is $12\ \text{cm}$ long and $3\ \text{cm}$ thick?
   a. 2
   b. 3
   c. 4
   d. 9

32. Which statement correctly describes the simple machines, like the crank in the image, that make up an Archimedes screw and the forces it applies?

   a. The crank is a wedge in which the IMA is the length of the tube divided by the radius of the tube. The applied force is the effort force and the weight of the water is the resistance force.
   b. The crank is an inclined plane in which the IMA is the length of the tube divided by the radius of the tube. The applied force is the effort force and the weight of the water is the resistance force.
   c. The crank is a wheel and axle. The effort force of the crank becomes the resistance force of the screw.
   d. The crank is a wheel and axle. The resistance force of the crank becomes the effort force of the screw.

33. Refer to the pulley system on right in the image. Assume this pulley system is an ideal machine.
How hard would you have to pull on the rope to lift a 120 N

load?

How many meters of rope would you have to pull out of the system to lift the load 1 m?

a. 480 N
4 m
b. 480 N
$\frac{1}{4}$ m
c. 30 N
4 m
d. 30 N
$\frac{1}{4}$ m

Short Answer

9.1 Work, Power, and the Work–Energy Theorem

34. Describe two ways in which doing work on an object can increase its mechanical energy.

a. Raising an object to a higher elevation does work as it increases its PE; increasing the speed of an object does work as it increases its KE.
b. Raising an object to a higher elevation does work as it increases its KE; increasing the speed of an object does work as it increases its PE.
c. Raising an object to a higher elevation does work as it increases its PE; decreasing the speed of an object does work as it increases its KE.
d. Raising an object to a higher elevation does work as it increases its KE; decreasing the speed of an object does work as it increases its PE.

35. True or false—While riding a bicycle up a gentle hill, it is fairly easy to increase your potential energy, but to increase your kinetic energy would make you feel exhausted.

a. True
b. False

36. Which statement best explains why running on a track with constant speed at 3 m/s is not work, but climbing a mountain at 1 m/s is work?

a. At constant speed, change in the kinetic energy is zero but climbing a mountain produces change in the potential energy.
b. At constant speed, change in the potential energy is zero, but climbing a mountain produces change in the kinetic energy.
c. At constant speed, change in the kinetic energy is finite, but climbing a mountain produces no change in the potential energy.
d. At constant speed, change in the potential energy is finite, but climbing a mountain produces no change in the kinetic energy.

37. You start at the top of a hill on a bicycle and coast to the bottom without applying the brakes. By the time you reach the bottom of the hill, work has been done on you and your bicycle, according to the equation: $W = \frac{1}{2}m(v_2^2 - v_1^2)$ If m is the mass of you and your bike, what are v_1 and v_2?

a. v_1 is your speed at the top of the hill, and v_2 is your speed at the bottom.
b. v_1 is your speed at the bottom of the hill, and v_2 is your speed at the top.
c. v_1 is your displacement at the top of the hill, and v_2 is your displacement at the bottom.
d. v_1 is your displacement at the bottom of the hill, and v_2 is your displacement at the top.

9.2 Mechanical Energy and Conservation of Energy

38. True or false—The formula for gravitational potential energy can be used to explain why joules, J, are equivalent to kg × mg² / s² . Show your work.

a. True
b. False

39. Which statement best explains why accelerating a car from 20 mph to 40 mph quadruples its kinetic energy?

a. Because kinetic energy is directly proportional to the square of the velocity.
b. Because kinetic energy is inversely proportional to the square of the velocity.
c. Because kinetic energy is directly proportional to the fourth power of the velocity.
d. Because kinetic energy is inversely proportional to

the fourth power of the velocity.

40. A coin falling through a vacuum loses no energy to friction, and yet, after it hits the ground, it has lost all its potential and kinetic energy. Which statement best explains why the law of conservation of energy is still valid in this case?

a. When the coin hits the ground, the ground gains potential energy that quickly changes to thermal energy.
b. When the coin hits the ground, the ground gains kinetic energy that quickly changes to thermal energy.
c. When the coin hits the ground, the ground gains thermal energy that quickly changes to kinetic energy.
d. When the coin hits the ground, the ground gains thermal energy that quickly changes to potential energy.

41. True or false—A marble rolls down a slope from height h_1 and up another slope to height h_2, where ($h_2 < h_1$). The difference $mg(h_1 - h_2)$ is equal to the heat lost due to the friction.

a. True
b. False

9.3 Simple Machines

42. Why would you expect the lever shown in the top image to have a greater efficiency than the inclined plane shown in the bottom image?

a. The resistance arm is shorter in case of the inclined plane.
b. The effort arm is shorter in case of the inclined plane.
c. The area of contact is greater in case of the inclined plane.

43. Why is the wheel on a wheelbarrow **not** a simple machine in the same sense as the simple machine in the image?

a. The wheel on the wheelbarrow has no fulcrum.
b. The center of the axle is not the fulcrum for the wheels of a wheelbarrow.
c. The wheelbarrow differs in the way in which load is attached to the axle.
d. The wheelbarrow has less resistance force than a wheel and axle design.

44. A worker pulls down on one end of the rope of a pulley system with a force of 75 N to raise a hay bale tied to the other end of the rope. If she pulls the rope down 2.0 m and the bale raises 1.0 m, what else would you have to know to calculate the efficiency of the pulley system?

a. the weight of the worker
b. the weight of the hay bale
c. the radius of the pulley
d. the height of the pulley from ground

45. True or false—A boy pushed a box with a weight of 300 N up a ramp. He said that, because the ramp was 1.0 m high and 3.0 m long, he must have been pushing with force of exactly 100 N.

a. True
b. False

Extended Response

9.1 Work, Power, and the Work–Energy Theorem

46. Work can be negative as well as positive because an object or system can do work on its surroundings as well as have work done on it. Which of the following statements describes:
a situation in which an object does work on its surroundings by decreasing its velocity and
a situation in which an object can do work on its surroundings by decreasing its altitude?

a. A gasoline engine burns less fuel at a slower speed. Solar cells capture sunlight to generate electricity.

b. A hybrid car charges its batteries as it decelerates.
Falling water turns a turbine to generate electricity.
c. Airplane flaps use air resistance to slow down for landing.
Rising steam turns a turbine to generate electricity.
d. An electric train requires less electrical energy as it decelerates.
A parachute captures air to slow a skydiver's fall.

47. A boy is pulling a girl in a child's wagon at a constant speed. He begins to pull harder, which increases the speed of the wagon. Which of the following describes two ways you could calculate the change in energy of the wagon and girl if you had all the information you needed?
a. Calculate work done from the force and the velocity.
Calculate work done from the change in the potential energy of the system.
b. Calculate work done from the force and the displacement.
Calculate work done from the change in the potential energy of the system.
c. Calculate work done from the force and the velocity.
Calculate work done from the change in the kinetic energy of the system.
d. Calculate work done from the force and the displacement.
Calculate work done from the change in the kinetic energy of the system.

9.2 Mechanical Energy and Conservation of Energy

48. Acceleration due to gravity on the moon is 1.6 m/s^2 or about 16% of the value of *g* on Earth.
If an astronaut on the moon threw a moon rock to a height of 7.8 m, what would be its velocity as it struck the moon's surface?
How would the fact that the moon has no atmosphere affect the velocity of the falling moon rock? Explain your answer.
a. The velocity of the rock as it hits the ground would be 5.0 m/s. Due to the lack of air friction, there would be complete transformation of the potential energy into the kinetic energy as the rock hits the moon's surface.
b. The velocity of the rock as it hits the ground would be 5.0 m/s. Due to the lack of air friction, there would be incomplete transformation of the potential energy into the kinetic energy as the rock hits the moon's surface.
c. The velocity of the rock as it hits the ground would be 12 m/s. Due to the lack of air friction, there would be complete transformation of the potential energy into the kinetic energy as the rock hits the moon's surface.
d. The velocity of the rock as it hits the ground would be 12 m/s. Due to the lack of air friction, there would be incomplete transformation of the potential energy into the kinetic energy as the rock hits the moon's surface.

49. A boulder rolls from the top of a mountain, travels across a valley below, and rolls part way up the ridge on the opposite side. Describe all the energy transformations taking place during these events and identify when they happen.
a. As the boulder rolls down the mountainside, KE is converted to PE. As the boulder rolls up the opposite slope, PE is converted to KE. The boulder rolls only partway up the ridge because some of the PE has been converted to thermal energy due to friction.
b. As the boulder rolls down the mountainside, KE is converted to PE. As the boulder rolls up the opposite slope, KE is converted to PE. The boulder rolls only partway up the ridge because some of the PE has been converted to thermal energy due to friction.
c. As the boulder rolls down the mountainside, PE is converted to KE. As the boulder rolls up the opposite slope, PE is converted to KE. The boulder rolls only partway up the ridge because some of the PE has been converted to thermal energy due to friction.
d. As the boulder rolls down the mountainside, PE is converted to KE. As the boulder rolls up the opposite slope, KE is converted to PE. The boulder rolls only partway up the ridge because some of the PE has been converted to thermal energy due to friction.

9.3 Simple Machines

50. To dig a hole, one holds the handles together and thrusts the blades of a posthole digger, like the one in the image, into the ground. Next, the handles are pulled apart, which squeezes the dirt between them, making it possible to remove the dirt from the hole. This complex machine is composed of two pairs of two different simple machines. Identify and describe the parts that are simple machines and explain how you would find the IMA of each type of simple machine.

a. Each handle and its attached blade is a lever with the

fulcrum at the hinge. Each blade is a wedge.
The IMA of a lever would be the length of the handle divided by the length of the blade. The IMA of the wedges would be the length of the blade divided by its width.

b. Each handle and its attached to blade is a lever with the fulcrum at the end. Each blade is a wedge.
The IMA of a lever would be the length of the handle divided by the length of the blade. The IMA of the wedges would be the length of the blade divided by its width.

c. Each handle and its attached blade is a lever with the fulcrum at the hinge. Each blade is a wedge.
The IMA of a lever would be the length of the handle multiplied by the length of the blade. The IMA of the wedges would be the length of the blade multiplied by its width.

d. Each handle and its attached blade is a lever with the fulcrum at the end. Each blade is a wedge.
The IMA of a lever would be the length of the handle multiplied by the length of the blade. The IMA of the wedges would be the length of the blade multiplied by its width.

51. A wooden crate is pulled up a ramp that is 1.0 m high and 6.0 m long. The crate is attached to a rope that is wound around an axle with a radius of 0.020 m . The axle is turned by a 0.20 m long handle. What is the overall IMA of the complex machine?

A. 6
B. 10
C. 16
D. 60

CHAPTER 10
Special Relativity

Figure 10.1 Special relativity explains why travel to other star systems, such as these in the Orion Nebula, is unlikely using our current level of technology. (s58y, Flickr)

Chapter Outline

10.1 Postulates of Special Relativity

10.2 Consequences of Special Relativity

INTRODUCTION Have you ever dreamed of traveling to other planets in faraway star systems? The trip might seem possible by traveling fast enough, but you will read in this chapter why it is not. In 1905, Albert Einstein developed the theory of **special relativity**. Einstein developed the theory to help explain inconsistencies between the equations describing electromagnetism and Newtonian mechanics, and to explain why the ether did not exist. This theory explains the limit on an object's speed among other implications.

Relativity is the study of how different observers moving with respect to one another measure the same events. Galileo and Newton developed the first correct version of classical relativity. Einstein developed the modern theory of relativity. Modern relativity is divided into two parts. Special relativity deals with observers moving at constant velocity. **General relativity** deals with observers moving at constant acceleration. Einstein's theories of relativity made revolutionary predictions. Most importantly, his predictions have been verified by experiments.

In this chapter, you learn how experiments and puzzling contradictions in existing theories led to the development of the theory of special relativity. You will also learn the simple postulates on which the theory was based; a postulate is a statement that is assumed to be true for the purposes of reasoning in a scientific or mathematic argument.

10.1 Postulates of Special Relativity

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Describe the experiments and scientific problems that led Albert Einstein to develop the special theory of relativity
- Understand the postulates on which the special theory of relativity was based

Section Key Terms

ether	frame of reference	inertial reference frame
general relativity	postulate	relativity
simultaneity	special relativity	

Scientific Experiments and Problems

Relativity is not new. Way back around the year 1600, Galileo explained that motion is relative. Wherever you happen to be, it seems like you are at a fixed point and that everything moves with respect to you. Everyone else feels the same way. Motion is always measured with respect to a fixed point. This is called establishing a **frame of reference**. But the choice of the point is arbitrary, and all frames of reference are equally valid. A passenger in a moving car is not moving with respect to the driver, but they are both moving from the point of view of a person on the sidewalk waiting for a bus. They are moving even faster as seen by a person in a car coming toward them. It is all relative.

TIPS FOR SUCCESS

A frame of reference is not a complicated concept. It is just something you decide is a fixed point or group of connected points. It is completely up to you. For example, when you look up at celestial objects in the sky, you choose the earth as your frame of reference, and the sun, moon, etc., seem to move across the sky.

Light is involved in the discussion of relativity because theories related to electromagnetism are inconsistent with Galileo's and Newton's explanation of relativity. The true nature of light was a hot topic of discussion and controversy in the late 19th century. At the time, it was not generally believed that light could travel across empty space. It was known to travel as waves, and all other types of energy that propagated as waves needed to travel though a material medium. It was believed that space was filled with an invisible medium that light waves traveled through. This imaginary (as it turned out) material was called the **ether** (also spelled aether). It was thought that everything moved through this mysterious fluid. In other words, ether was the one fixed frame of reference. The Michelson–Morley experiment proved it was not.

In 1887, Albert Michelson and Edward Morley designed the interferometer shown in Figure 10.2 to measure the speed of Earth through the ether. A light beam is split into two perpendicular paths and then recombined. Recombining the waves produces an inference pattern, with a bright fringe at the locations where the two waves arrive in phase; that is, with the crests of both waves arriving together and the troughs arriving together. A dark fringe appears where the crest of one wave coincides with a trough of the other, so that the two cancel. If Earth is traveling through the ether as it orbits the sun, the peaks in one arm would take longer than in the other to reach the same location. The places where the two waves arrive in phase would change, and the interference pattern would shift. But, using the interferometer, there was no shift seen! This result led to two conclusions: that there is no ether and that the speed of light is the same regardless of the relative motion of source and observer. The Michelson–Morley investigation has been called the most famous failed experiment in history.

Figure 10.2 This is a diagram of the instrument used in the Michelson–Morley experiment.

To see what Michelson and Morley expected to find when they measured the speed of light in two directions, watch this animation (http://openstax.org/l/28MMexperiment) . In the video, two people swimming in a lake are represented as an analogy to light beams leaving Earth as it moves through the ether (if there were any ether). The swimmers swim away from and back to a platform that is moving through the water. The swimmers swim in different directions with respect to the motion of the platform. Even though they swim equal distances at the same speed, the motion of the platform causes them to arrive at different times.

Einstein's Postulates

The results described above left physicists with some puzzling and unsettling questions such as, why doesn't light emitted by a fast-moving object travel faster than light from a street lamp? A radical new theory was needed, and Albert Einstein, shown in Figure 10.3, was about to become everyone's favorite genius. Einstein began with two simple **postulates** based on the two things we have discussed so far in this chapter.

1. The laws of physics are the same in all inertial reference frames.
2. The speed of light is the same in all inertial reference frames and is not affected by the speed of its source.

Figure 10.3 Albert Einstein (1879–1955) developed modern relativity and also made fundamental contributions to the foundations of quantum mechanics. (The Library of Congress)

The speed of light is given the symbol *c* and is equal to exactly 299,792,458 m/s. This is the speed of light in vacuum; that is, in the absence of air. For most purposes, we round this number off to 3.00×10^8 m/s The term **inertial reference frame** simply

refers to a frame of reference where all objects follow Newton's first law of motion: Objects at rest remain at rest, and objects in motion remain in motion at a constant velocity in a straight line, unless acted upon by an external force. The inside of a car moving along a road at constant velocity and the inside of a stationary house are inertial reference frames.

WATCH PHYSICS

The Speed of Light

This lecture on light summarizes the most important facts about the speed of light. If you are interested, you can watch the whole video, but the parts relevant to this chapter are found between 3:25 and 5:10, which you find by running your cursor along the bottom of the video.

Click to view content (https://www.youtube.com/embed/rLNM8zI4Q_M)

GRASP CHECK

An airliner traveling at 200 m/s emits light from the front of the plane. Which statement describes the speed of the light?

a. It travels at a speed of $c + 200$ m/s.
b. It travels at a speed of $c - 200$ m/s.
c. It travels at a speed c, like all light.
d. It travels at a speed slightly less than c.

Snap Lab

Measure the Speed of Light

In this experiment, you will measure the speed of light using a microwave oven and a slice of bread. The waves generated by a microwave oven are not part of the visible spectrum, but they are still electromagnetic radiation, so they travel at the speed of light. If we know the wavelength, λ, and frequency, f, of a wave, we can calculate its speed, v, using the equation $v = \lambda f$. You can measure the wavelength. You will find the frequency on a label on the back of a microwave oven. The wave in a microwave is a standing wave with areas of high and low intensity. The high intensity sections are one-half wavelength apart.

- High temperature: Very hot temperatures are encountered in this lab. These can cause burns.

- a microwave oven
- one slice of plain white bread
- a centimeter ruler
- a calculator

1. Work with a partner.
2. Turn off the revolving feature of the microwave oven or remove the wheels under the microwave dish that make it turn. It is important that the dish does not turn.
3. Place the slice of bread on the dish, set the microwave on high, close the door, run the microwave for about 15 seconds.
4. A row of brown or black marks should appear on the bread. Stop the microwave as soon as they appear. Measure the distance between two adjacent burn marks and multiply the result by 2. This is the wavelength.
5. The frequency of the waves is written on the back of the microwave. Look for something like "2,450 MHz." Hz is the unit hertz, which means *per second*. The M represents mega, which stands for million, so multiply the number by 10^6.
6. Express the wavelength in meters and multiply it times the frequency. If you did everything correctly, you will get a number very close to the speed of light. Do not eat the bread. It is a general laboratory safety rule never to eat anything in the lab.

GRASP CHECK

How does your measured value of the speed of light compare to the accepted value (% error)?

a. The measured value of speed will be equal to *c*.
b. The measured value of speed will be slightly less than *c*.
c. The measured value of speed will be slightly greater than *c*.
d. The measured value of speed will depend on the frequency of the microwave.

Einstein's postulates were carefully chosen, and they both seemed very likely to be true. Einstein proceeded despite realizing that these two ideas taken together and applied to extreme conditions led to results that contradict Newtonian mechanics. He just took the ball and ran with it.

In the traditional view, velocities are additive. If you are running at 3 m/s and you throw a ball forward at a speed of 10 m/s, the ball should have a net speed of 13 m/s. However, according to relativity theory, the speed of a moving light source is not added to the speed of the emitted light.

In addition, Einstein's theory shows that if you were moving forward relative to Earth at nearly *c* (the speed of light) and could throw a ball forward at *c*, an observer at rest on the earth would not see the ball moving at nearly twice the speed of light. The observer would see it moving at a speed that is still less than *c*. This result conforms to both of Einstein's postulates: The speed of light has a fixed maximum and neither reference frame is privileged.

Consider how we measure elapsed time. If we use a stopwatch, for example, how do we know when to start and stop the watch? One method is to use the arrival of light from the event, such as observing a light turn green to start a drag race. The timing will be more accurate if some sort of electronic detection is used, avoiding human reaction times and other complications.

Now suppose we use this method to measure the time interval between two flashes of light produced by flash lamps on a moving train. (See Figure 10.4)

Figure 10.4 Light arriving to observer A as seen by two different observers.

A woman (observer A) is seated in the center of a rail car, with two flash lamps at opposite sides equidistant from her. Multiple light rays that are emitted from the flash lamps move towards observer A, as shown with arrows. A velocity vector arrow for the rail car is shown towards the right. A man (observer B) standing on the platform is facing the woman and also observes the flashes of light.

Observer A moves with the lamps on the rail car as the rail car moves towards the right of observer B. Observer B receives the light flashes simultaneously, and sees the bulbs as both having flashed at the same time. However, he sees observer A receive the flash from the right first. Because the pulse from the right reaches her first, in her frame of reference she sees the bulbs as not having flashed simultaneously. Here, a relative velocity between observers affects whether two events at well-separated locations are observed to be simultaneous. **Simultaneity**, or whether different events occur at the same instant, depends on the frame of reference of the observer. Remember that velocity equals distance divided by time, so $t = d/\mathbf{v}$. If velocity appears to be different, then duration of time appears to be different.

This illustrates the power of clear thinking. We might have guessed incorrectly that, if light is emitted simultaneously, then two observers halfway between the sources would see the flashes simultaneously. But careful analysis shows this not to be the case. Einstein was brilliant at this type of thought experiment (in German, *Gedankenexperiment*). He very carefully considered how an observation is made and disregarded what might seem obvious. The validity of thought experiments, of course, is determined by actual observation. The genius of Einstein is evidenced by the fact that experiments have repeatedly confirmed his theory of relativity. No experiments after that of Michelson and Morley were able to detect any ether medium. We will describe later how experiments also confirmed other predictions of **special relativity**, such as the distance between two objects and the time interval of two events being different for two observers moving with respect to each other.

In summary: Two events are defined to be simultaneous if an observer measures them as occurring at the same time (such as by receiving light from the events). Two events are not necessarily simultaneous to all observers.

The discrepancies between Newtonian mechanics and relativity theory illustrate an important point about how science advances. Einstein's theory did not replace Newton's but rather extended it. It is not unusual that a new theory must be developed to account for new information. In most cases, the new theory is built on the foundation of older theory. It is rare that old theories are completely replaced.

In this chapter, you will learn about the theory of special relativity, but, as mentioned in the introduction, Einstein developed two relativity theories: special and general. Table 10.1 summarizes the differences between the two theories.

Special Relativity	General Relativity
Published in 1905	Final form published in 1916
A theory of space-time	A theory of gravity
Applies to observers moving at constant speed	Applies to observers that are accelerating
Most useful in the field of nuclear physics	Most useful in the field of astrophysics
Accepted quickly and put to practical use by nuclear physicists and quantum chemists	Largely ignored until 1960 when new mathematical techniques made the theory more accessible and astronomers found some important applications
Also note that the theory of general relativity includes the theory of special relativity.	

Table 10.1 Comparing Special Relativity and General Relativity

WORKED EXAMPLE

Calculating the Time it Takes Light to Travel a Given Distance

The sun is 1.50×10^8 km from Earth. How long does it take light to travel from the sun to Earth in minutes and seconds?

Strategy

Identify knowns.

$$\text{Distance} = 1.50 \times 10^8 \text{ km}$$
$$\text{Speed} = 3.00 \times 10^8 \text{ km} \tag{10.1}$$

Identify unknowns.

Time

Find the equation that relates knowns and unknowns.

$$v = \frac{d}{t}; \quad t = \frac{d}{v} \tag{10.2}$$

Be sure to use consistent units.

Solution

$$t = \frac{d}{v} = \frac{(1.50 \times 10^8 km) \times \frac{10^3 \text{m}}{\text{km}}}{3.00 \times 10^8 \frac{\text{m}}{\text{s}}} = 5.00 \times 10^2 s$$
$$500 \text{ s} = 8 \text{ min and } 20 \text{ s}$$

Discussion

The answer is written as 5.00×10^2 rather than 500 in order to show that there are three significant figures. When astronomers witness an event on the sun, such as a sunspot, it actually happened minutes earlier. Compare 8 light *minutes* to the distance to stars, which are light *years* away. Any events on other stars happened years ago.

Practice Problems

1. Light travels through 1.00 m of water in 4.42×10^{-9} s. What is the speed of light in water?
   a. 4.42×10^{-9} m/s
   b. 4.42×10^{9} m/s
   c. 2.26×10^{8} m/s
   d. 226×10^{8} m/s

2. An astronaut on the moon receives a message from mission control on Earth. The signal is sent by a form of electromagnetic radiation and takes 1.28 s to travel the distance between Earth and the moon. What is the distance from Earth to the moon?
   a. 2.34×10^{5} km
   b. 2.34×10^{8} km
   c. 3.84×10^{5} km
   d. 3.84×10^{8} km

Check Your Understanding

3. Explain what is meant by a frame of reference.
   a. A frame of reference is a graph plotted between distance and time.
   b. A frame of reference is a graph plotted between speed and time.
   c. A frame of reference is the velocity of an object through empty space without regard to its surroundings.
   d. A frame of reference is an arbitrarily fixed point with respect to which motion of other points is measured.

4. Two people swim away from a raft that is floating downstream. One swims upstream and returns, and the other swims across the current and back. If this scenario represents the Michelson–Morley experiment, what do (i) the water, (ii) the swimmers, and (iii) the raft represent?
   a. the ether rays of light Earth
   b. rays of light the ether Earth
   c. the ether Earth rays of light
   d. Earth rays of light the ether

5. If Michelson and Morley had observed the interference pattern shift in their interferometer, what would that have indicated?
   a. The speed of light is the same in all frames of reference.
   b. The speed of light depends on the motion relative to the ether.
   c. The speed of light changes upon reflection from a surface.
   d. The speed of light in vacuum is less than 3.00×10^{8} m/s.

6. If you designate a point as being fixed and use that point to measure the motion of surrounding objects, what is the point called?
   a. An origin
   b. A frame of reference
   c. A moving frame
   d. A coordinate system

10.2 Consequences of Special Relativity

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Describe the relativistic effects seen in time dilation, length contraction, and conservation of relativistic momentum
- Explain and perform calculations involving mass-energy equivalence

Section Key Terms

binding energy	length contraction	mass defect	time dilation

proper length	relativistic	relativistic momentum
relativistic energy	relativistic factor	rest mass

Relativistic Effects on Time, Distance, and Momentum

Consideration of the measurement of elapsed time and simultaneity leads to an important relativistic effect. **Time dilation** is the phenomenon of time passing more slowly for an observer who is moving relative to another observer.

For example, suppose an astronaut measures the time it takes for light to travel from the light source, cross her ship, bounce off a mirror, and return. (See Figure 10.5.) How does the elapsed time the astronaut measures compare with the elapsed time measured for the same event by a person on the earth? Asking this question (another thought experiment) produces a profound result. We find that the elapsed time for a process depends on who is measuring it. In this case, the time measured by the astronaut is smaller than the time measured by the earth bound observer. The passage of time is different for the two observers because the distance the light travels in the astronaut's frame is smaller than in the earth bound frame. Light travels at the same speed in each frame, and so it will take longer to travel the greater distance in the earth bound frame.

Figure 10.5 (a) An astronaut measures the time Δt_0 for light to cross her ship using an electronic timer. Light travels a distance $2D$ in the astronaut's frame. (b) A person on the earth sees the light follow the longer path $2s$ and take a longer time Δt.

The relationship between Δt and Δt_0 is given by

$$\Delta t = \gamma \Delta t_0,$$

where γ is the **relativistic factor** given by

$$\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}},$$

and v and c are the speeds of the moving observer and light, respectively.

TIPS FOR SUCCESS

Try putting some values for v into the expression for the relativistic factor (γ). Observe at which speeds this factor will make a difference and when γ is so close to 1 that it can be ignored. Try 225 m/s, the speed of an airliner; 2.98×10^4 m/s, the speed of Earth in its orbit; and 2.990×10^8 m/s, the speed of a particle in an accelerator.

Notice that when the velocity v is small compared to the speed of light c, then v/c becomes small, and γ becomes close to 1. When this happens, time measurements are the same in both frames of reference. **Relativistic** effects, meaning those that have to do with special relativity, usually become significant when speeds become comparable to the speed of light. This is seen to be the case for time dilation.

You may have seen science fiction movies in which space travelers return to Earth after a long trip to find that the planet and everyone on it has aged much more than they have. This type of scenario is a based on a thought experiment, known as the twin paradox, which imagines a pair of twins, one of whom goes on a trip into space while the other stays home. When the space traveler returns, she finds her twin has aged much more than she. This happens because the traveling twin has been in two frames of reference, one leaving Earth and one returning.

Time dilation has been confirmed by comparing the time recorded by an atomic clock sent into orbit to the time recorded by a clock that remained on Earth. GPS satellites must also be adjusted to compensate for time dilation in order to give accurate positioning.

Have you ever driven on a road, like that shown in Figure 10.6, that seems like it goes on forever? If you look ahead, you might say you have about 10 km left to go. Another traveler might say the road ahead looks like it is about 15 km long. If you both measured the road, however, you would agree. Traveling at everyday speeds, the distance you both measure would be the same. You will read in this section, however, that this is not true at relativistic speeds. Close to the speed of light, distances measured are not the same when measured by different observers moving with respect to one other.

Figure 10.6 People might describe distances differently, but at relativistic speeds, the distances really are different. (Corey Leopold, Flickr)

One thing all observers agree upon is their relative speed. When one observer is traveling away from another, they both see the other receding at the same speed, regardless of whose frame of reference is chosen. Remember that speed equals distance divided by time: $v = d/t$. If the observers experience a difference in elapsed time, they must also observe a difference in distance traversed. This is because the ratio d/t must be the same for both observers.

The shortening of distance experienced by an observer moving with respect to the points whose distance apart is measured is called **length contraction**. **Proper length**, L_0, is the distance between two points measured in the reference frame where the observer and the points are at rest. The observer in motion with respect to the points measures L. These two lengths are related by the equation

$$L = \frac{L_0}{\gamma}.$$

Because γ is the same expression used in the time dilation equation above, the equation becomes

$$L = L_0\sqrt{1 - \frac{v^2}{c^2}}.$$

To see how length contraction is seen by a moving observer, go to this simulation (http://openstax.org/l/28simultaneity) . Here you can also see that simultaneity, time dilation, and length contraction are interrelated phenomena.

This link is to a simulation that illustrates the relativity of simultaneous events.

In classical physics, momentum is a simple product of mass and velocity. When special relativity is taken into account, objects that have mass have a speed limit. What effect do you think mass and velocity have on the momentum of objects moving at relativistic speeds; i.e., speeds close to the speed of light?

Momentum is one of the most important concepts in physics. The broadest form of Newton's second law is stated in terms of momentum. Momentum is conserved in classical mechanics whenever the net external force on a system is zero. This makes momentum conservation a fundamental tool for analyzing collisions. We will see that momentum has the same importance in modern physics. **Relativistic momentum** is conserved, and much of what we know about subatomic structure comes from the analysis of collisions of accelerator-produced relativistic particles.

One of the postulates of special relativity states that the laws of physics are the same in all inertial frames. Does the law of conservation of momentum survive this requirement at high velocities? The answer is yes, provided that the momentum is defined as follows.

Relativistic momentum, $\mathbf{p}$, is classical momentum multiplied by the relativistic factor γ.

$$\mathbf{p} = \gamma m\mathbf{u}, \tag{10.3}$$

where m is the **rest mass** of the object (that is, the mass measured at rest, without any γ factor involved), $\mathbf{u}$ is its velocity relative to an observer, and γ, as before, is the relativistic factor. We use the mass of the object as measured at rest because we cannot determine its mass while it is moving.

Note that we use $\mathbf{u}$ for velocity here to distinguish it from relative velocity $\mathbf{v}$ between observers. Only one observer is being considered here. With $\mathbf{p}$ defined in this way, $\mathbf{p}_{\text{tot}}$ is conserved whenever the net external force is zero, just as in classical physics. Again we see that the relativistic quantity becomes virtually the same as the classical at low velocities. That is, relativistic momentum $\gamma m\mathbf{u}$ becomes the classical $m\mathbf{u}$ at low velocities, because γ is very nearly equal to 1 at low velocities.

Relativistic momentum has the same intuitive feel as classical momentum. It is greatest for large masses moving at high velocities. Because of the factor γ, however, relativistic momentum behaves differently from classical momentum by approaching infinity as $\mathbf{u}$ approaches c. (See Figure 10.7.) This is another indication that an object with mass cannot reach the speed of light. If it did, its momentum would become infinite, which is an unreasonable value.

Figure 10.7 Relativistic momentum approaches infinity as the velocity of an object approaches the speed of light.

Relativistic momentum is defined in such a way that the conservation of momentum will hold in all inertial frames. Whenever the net external force on a system is zero, relativistic momentum is conserved, just as is the case for classical momentum. This

has been verified in numerous experiments.

Mass-Energy Equivalence

Let us summarize the calculation of relativistic effects on objects moving at speeds near the speed of light. In each case we will need to calculate the relativistic factor, given by

$$\gamma = \frac{1}{\sqrt{1 - \frac{\mathbf{v}^2}{c^2}}},$$

where **v** and c are as defined earlier. We use **u** as the velocity of a particle or an object in one frame of reference, and **v** for the velocity of one frame of reference with respect to another.

Time Dilation

Elapsed time on a moving object, Δt_0, as seen by a stationary observer is given by $\Delta t = \gamma \Delta t_0$, where Δt_0 is the time observed on the moving object when it is taken to be the frame or reference.

Length Contraction

Length measured by a person at rest with respect to a moving object, L, is given by

$$L = \frac{L_0}{\gamma},$$

where L_0 is the length measured on the moving object.

Relativistic Momentum

Momentum, $\mathbf{p}$, of an object of mass, m, traveling at relativistic speeds is given by $\mathbf{p} = \gamma m \mathbf{u}$, where **u** is velocity of a moving object as seen by a stationary observer.

Relativistic Energy

The original source of all the energy we use is the conversion of mass into energy. Most of this energy is generated by nuclear reactions in the sun and radiated to Earth in the form of electromagnetic radiation, where it is then transformed into all the forms with which we are familiar. The remaining energy from nuclear reactions is produced in nuclear power plants and in Earth's interior. In each of these cases, the source of the energy is the conversion of a small amount of mass into a large amount of energy. These sources are shown in Figure 10.8.

Figure 10.8 The sun (a) and the Susquehanna Steam Electric Station (b) both convert mass into energy. ((a) NASA/Goddard Space Flight Center, Scientific Visualization Studio; (b) U.S. government)

The first postulate of relativity states that the laws of physics are the same in all inertial frames. Einstein showed that the law of conservation of energy is valid relativistically, if we define energy to include a relativistic factor. The result of his analysis is that a particle or object of mass m moving at velocity **u** has **relativistic energy** given by

$$E = \gamma mc^2.$$

This is the expression for the total energy of an object of mass m at any speed **u** and includes both kinetic and potential energy. Look back at the equation for γ and you will see that it is equal to 1 when **u** is 0; that is, when an object is at rest. Then the rest

energy, E_0, is simply

$$E_0 = mc^2.$$

This is the correct form of Einstein's famous equation.

This equation is very useful to nuclear physicists because it can be used to calculate the energy released by a nuclear reaction. This is done simply by subtracting the mass of the products of such a reaction from the mass of the reactants. The difference is the m in $E_0 = mc^2$. Here is a simple example:

A positron is a type of antimatter that is just like an electron, except that it has a positive charge. When a positron and an electron collide, their masses are completely annihilated and converted to energy in the form of gamma rays. Because both particles have a rest mass of 9.11×10^{-31} kg, we multiply the mc^2 term by 2. So the energy of the gamma rays is

$$\begin{aligned} E_0 &= 2(9.11 \times 10^{-31}\ \text{kg})\left(3.00 \times 10^8\ \tfrac{\text{m}}{\text{s}}\right)^2 \\ &= 1.64 \times 10^{-13}\ \tfrac{\text{kg}\cdot\text{m}^2}{\text{s}^2} \\ &= 1.64 \times 10^{-13}\,\text{J} \end{aligned} \tag{10.4}$$

where we have the expression for the joule (J) in terms of its SI base units of kg, m, and s. In general, the nuclei of stable isotopes have less mass then their constituent subatomic particles. The energy equivalent of this difference is called the **binding energy** of the nucleus. This energy is released during the formation of the isotope from its constituent particles because the product is more stable than the reactants. Expressed as mass, it is called the **mass defect**. For example, a helium nucleus is made of two neutrons and two protons and has a mass of 4.0003 atomic mass units (u). The sum of the masses of two protons and two neutrons is 4.0330 u. The mass defect then is 0.0327 u. Converted to kg, the mass defect is 5.0442×10^{-30} kg. Multiplying this mass times c^2 gives a binding energy of 4.540×10^{-12} J. This does not sound like much because it is only one atom. If you were to make one gram of helium out of neutrons and protons, it would release 683,000,000,000 J. By comparison, burning one gram of coal releases about 24 J.

BOUNDLESS PHYSICS

The RHIC Collider

Figure 10.9 shows the Brookhaven National Laboratory in Upton, NY. The circular structure houses a particle accelerator called the RHIC, which stands for Relativistic Heavy Ion Collider. The heavy ions in the name are gold nuclei that have been stripped of their electrons. Streams of ions are accelerated in several stages before entering the big ring seen in the figure. Here, they are accelerated to their final speed, which is about 99.7 percent the speed of light. Such high speeds are called relativistic. All the relativistic phenomena we have been discussing in this chapter are very pronounced in this case. At this speed $\gamma = 12.9$, so that relativistic time dilates by a factor of about 13, and relativistic length contracts by the same factor.

Figure 10.9 Brookhaven National Laboratory. The circular structure houses the RHIC. (energy.gov, Wikimedia Commons)

Two ion beams circle the 2.4-mile long track around the big ring in opposite directions. The paths can then be made to cross, thereby causing ions to collide. The collision event is very short-lived but amazingly intense. The temperatures and pressures produced are greater than those in the hottest suns. At 4 trillion degrees Celsius, this is the hottest material ever created in a

laboratory

But what is the point of creating such an extreme event? Under these conditions, the neutrons and protons that make up the gold nuclei are smashed apart into their components, which are called quarks and gluons. The goal is to recreate the conditions that theorists believe existed at the very beginning of the universe. It is thought that, at that time, matter was a sort of soup of quarks and gluons. When things cooled down after the initial bang, these particles condensed to form protons and neutrons.

Some of the results have been surprising and unexpected. It was thought the quark-gluon soup would resemble a gas or plasma. Instead, it behaves more like a liquid. It has been called a *perfect* liquid because it has virtually no viscosity, meaning that it has no resistance to flow.

GRASP CHECK

Calculate the relativistic factor γ, for a particle traveling at 99.7 percent of the speed of light.

a. 0.08
b. 0.71
c. 1.41
d. 12.9

WORKED EXAMPLE

The Speed of Light

One night you are out looking up at the stars and an extraterrestrial spaceship flashes across the sky. The ship is 50 meters long and is travelling at 95 percent of the speed of light. What would the ship's length be when measured from your earthbound frame of reference?

Strategy

List the knowns and unknowns.

Knowns: proper length of the ship, $L_o = 50$ m; velocity, $\mathbf{v}, = 0.95c$

Unknowns: observed length of the ship accounting for relativistic length contraction, L.

Choose the relevant equation.

$$L = \frac{L_0}{\gamma} = L_0\sqrt{1 - \frac{\mathbf{u}^2}{c^2}}$$

Solution

$$L = 50\text{ m}\sqrt{1 - \frac{(0.95)^2 c^2}{c^2}} = 50\text{ m}\sqrt{1 - (0.95)^2} = 16\ \text{ m}$$

Discussion

Calculations of γ can usually be simplified in this way when v is expressed as a percentage of c because the c^2 terms cancel. Be sure to also square the decimal representing the percentage before subtracting from 1. Note that the aliens will still see the length as L_o because they are moving with the frame of reference that is the ship.

Practice Problems

7. Calculate the relativistic factor, γ, for an object traveling at 2.00×10^8 m/s.
   a. 0.74
   b. 0.83
   c. 1.2
   d. 1.34

8. The distance between two points, called the proper length, Lo, is 1.00 km. An observer in motion with respect to the frame of

reference of the two points measures 0.800 km, which is L. What is the relative speed of the frame of reference with respect to the observer?

a. 1.80×10^8 m/s
b. 2.34×10^8 m/s
c. 3.84×10^8 m/s
d. 5.00×10^8 m/s

9. Consider the nuclear fission reaction $n + {}^{235}_{92}U \rightarrow {}^{137}_{55}Cs + {}^{97}_{37}Rb + 2n + E$. If a neutron has a rest mass of 1.009u, ${}^{235}_{92}U$ has a rest mass of 235.044u, ${}^{137}_{55}Cs$ has rest mass of 136.907u, and ${}^{97}_{37}Rb$ has a rest mass of 96.937u, what is the value of E in joules?

a. 1.8×10^{-11} J
b. 2.9×10^{-11} J
c. 1.8×10^{-10} J
d. 2.9×10^{-10} J

Solution

The correct answer is (b). The mass deficit in the reaction is $235.044\text{ u} - (136.907 + 96.937 + 1.009)\text{ u}$, or 0.191u. Converting that mass to kg and applying $E = mc^2$ to find the energy equivalent of the mass deficit gives

$$(0.191\text{ u})\left(1.66 \times 10^{-27}\text{kg/u}\right)\left(3.00 \times 10^8\text{m/s}\right)^2 \cong 2.85 \times 10^{-11}\text{J}.$$

10. Consider the nuclear fusion reaction ${}^2_1H + {}^2_1H \rightarrow {}^3_1H + {}^1_1H + E$. If ${}^2_1H$ has a rest mass of 2.014u, ${}^3_1H$ has a rest mass of 3.016u, and ${}^1_1H$ has a rest mass of 1.008u, what is the value of E in joules?

a. 6×10^{-13} J
b. 6×10^{-12} J
c. 6×10^{-11} J
d. 6×10^{-10} J

Solution

The correct answer is (a). The mass deficit in the reaction is $2\,(2.014\text{ u}) - (3.016 + 1.008)\text{ u}$, or 0.004u. Converting that mass to kg and applying $E = mc^2$ to find the energy equivalent of the mass deficit gives

$$(0.004\text{ u})\left(1.66 \times 10^{-27}\text{kg/u}\right)\left(3.00 \times 10^8\text{m/s}\right)^2 \cong 5.98 \times 10^{-13}\text{J}.$$

Check Your Understanding

11. Describe time dilation and state under what conditions it becomes significant.

a. When the speed of one frame of reference past another reaches the speed of light, a time interval between two events at the same location in one frame appears longer when measured from the second frame.
b. When the speed of one frame of reference past another becomes comparable to the speed of light, a time interval between two events at the same location in one frame appears longer when measured from the second frame.
c. When the speed of one frame of reference past another reaches the speed of light, a time interval between two events at the same location in one frame appears shorter when measured from the second frame.
d. When the speed of one frame of reference past another becomes comparable to the speed of light, a time interval between two events at the same location in one frame appears shorter when measured from the second frame.

12. The equation used to calculate relativistic momentum is $p = \gamma \cdot m \cdot u$. Define the terms to the right of the equal sign and state how m and u are measured.

a. γ is the relativistic factor, m is the rest mass measured when the object is at rest in the frame of reference, and u is the velocity of the frame.
b. γ is the relativistic factor, m is the rest mass measured when the object is at rest in the frame of reference, and u is the velocity relative to an observer.
c. γ is the relativistic factor, m is the relativistic mass $\left(\text{i.e.,}\frac{m}{\sqrt{1-\frac{u^2}{c^2}}}\right)$ measured when the object is moving in the frame of reference, and u is the velocity of the frame.

d. γ is the relativistic factor, m is the relativistic mass $\left(\text{i.e.,} \frac{m}{\sqrt{1-\frac{u^2}{c^2}}}\right)$ measured when the object is moving in the frame of reference, and u is the velocity relative to an observer.

13. Describe length contraction and state when it occurs.

a. When the speed of an object becomes the speed of light, its length appears to shorten when viewed by a stationary observer.

b. When the speed of an object approaches the speed of light, its length appears to shorten when viewed by a stationary observer.

c. When the speed of an object becomes the speed of light, its length appears to increase when viewed by a stationary observer.

d. When the speed of an object approaches the speed of light, its length appears to increase when viewed by a stationary observer.

KEY TERMS

binding energy the energy equivalent of the difference between the mass of a nucleus and the masses of its nucleons

ether scientists once believed there was a medium that carried light waves; eventually, experiments proved that ether does not exist

frame of reference the point or collection of points arbitrarily chosen, which motion is measured in relation to

general relativity the theory proposed to explain gravity and acceleration

inertial reference frame a frame of reference where all objects follow Newton's first law of motion

length contraction the shortening of an object as seen by an observer who is moving relative to the frame of reference of the object

mass defect the difference between the mass of a nucleus and the masses of its nucleons

postulate a statement that is assumed to be true for the purposes of reasoning in a scientific or mathematic argument

proper length the length of an object within its own frame of reference, as opposed to the length observed by an observer moving relative to that frame of reference

relativistic having to do with modern relativity, such as the effects that become significant only when an object is moving close enough to the speed of light for γ to be significantly greater than 1

relativistic energy the total energy of a moving object or particle $E = \gamma mc^2$, which includes both its rest energy mc^2 and its kinetic energy

relativistic factor $\gamma = \frac{1}{\sqrt{1-\frac{\mathbf{u}^2}{c^2}}}$, where **u** is the velocity of a moving object and c is the speed of light

relativistic momentum $\mathbf{p} = \gamma m\mathbf{u}$, where γ is the relativistic factor, m is rest mass of an object, and **u** is the velocity relative to an observer

relativity the explanation of how objects move relative to one another

rest mass the mass of an object that is motionless with respect to its frame of reference

simultaneity the property of events that occur at the same time

special relativity the theory proposed to explain the consequences of requiring the speed of light and the laws of physics to be the same in all inertial frames

time dilation the contraction of time as seen by an observer in a frame of reference that is moving relative to the observer

SECTION SUMMARY

10.1 Postulates of Special Relativity

- One postulate of special relativity theory is that the laws of physics are the same in all inertial frames of reference.
- The other postulate is that the speed of light in a vacuum is the same in all inertial frames.
- Einstein showed that simultaneity, or lack of it, depends on the frame of reference of the observer.

10.2 Consequences of Special Relativity

- Time dilates, length contracts, and momentum increases as an object approaches the speed of light.
- Energy and mass are interchangeable, according to the relationship $E = mc2$. The laws of conservation of mass and energy are combined into the law of conservation of mass-energy.

KEY EQUATIONS

10.1 Postulates of Special Relativity

speed of light	$v = \lambda f$
constant value for the speed of light	$c = 3.00 \times 10^8 \text{ m/s}$

10.2 Consequences of Special Relativity

elapsed time	$\Delta t = \gamma \Delta t_0$
relativistic factor	$\gamma = \frac{1}{\sqrt{1-\frac{\mathbf{u}^2}{c^2}}}$
length contraction	$L = \frac{L_0}{\gamma}$
relativistic momentum	$\mathbf{p} = \gamma m\mathbf{u}$

relativistic energy $E = \gamma mc^2$

rest energy $E_0 = mc^2$

CHAPTER REVIEW

Concept Items

10.1 Postulates of Special Relativity

1. Why was it once believed that light must travel through a medium and could not propagate across empty space?
   a. The longitudinal nature of light waves implies this.
   b. Light shows the phenomenon of diffraction.
   c. The speed of light is the maximum possible speed.
   d. All other wave energy needs a medium to travel.

2. Describe the relative motion of Earth and the sun:
   1. if Earth is taken as the inertial frame of reference and
   2. if the sun is taken as the inertial frame of reference.
      a. 1. Earth is at rest and the sun orbits Earth.
         2. The sun is at rest and Earth orbits the sun.
      b. 1. The sun is at rest and Earth orbits the sun.
         2. Earth is at rest and the sun orbits Earth.
      c. 1. The sun is at rest and Earth orbits the sun.
         2. The sun is at rest and Earth orbits the sun.
      d. 1. Earth is at rest and the sun orbits Earth.
         2. Earth is at rest and the sun orbits Earth.

10.2 Consequences of Special Relativity

3. A β particle (a free electron) is speeding around the track in a cyclotron, rapidly gaining speed. How will the particle's momentum change as its speed approaches the speed of light? Explain.
   a. The particle's momentum will rapidly decrease.
   b. The particle's momentum will rapidly increase.
   c. The particle's momentum will remain constant.
   d. The particle's momentum will approach zero.

4. An astronaut goes on a long space voyage at near the speed of light. When she returns home, how will her age compare to the age of her twin who stayed on Earth?
   a. Both of them will be the same age.
   b. This is a paradox and hence the ages cannot be compared.
   c. The age of the twin who traveled will be less than the age of her twin.
   d. The age of the twin who traveled will be greater than the age of her twin.

5. A comet reaches its greatest speed as it travels near the sun. True or false— Relativistic effects make the comet's tail look longer to an observer on Earth.
   a. True
   b. False

Critical Thinking Items

10.1 Postulates of Special Relativity

6. Explain how the two postulates of Einstein's theory of special relativity, when taken together, could lead to a situation that seems to contradict the mechanics and laws of motion as described by Newton.
   a. In Newtonian mechanics, velocities are multiplicative but the speed of a moving light source cannot be multiplied to the speed of light because, according to special relativity, the speed of light is the maximum speed possible.
   b. In Newtonian mechanics, velocities are additive but the speed of a moving light source cannot be added to the speed of light because the speed of light is the maximum speed possible.
   c. An object that is at rest in one frame of reference may appear to be in motion in another frame of reference, while in Newtonian mechanics such a situation is not possible.
   d. The postulates of Einstein's theory of special relativity do not contradict any situation that Newtonian mechanics explains.

7. It takes light 6.0 minutes to travel from the sun to the planet Venus. How far is Venus from the sun?
   a. 18×10^6 km
   b. 18×10^8 km
   c. 1.08×10^{11} km
   d. 1.08×10^8 km

8. In 2003, Earth and Mars were the closest they had been in 50,000 years. The two planets were aligned so that Earth was between Mars and the sun. At that time it took light from the sun 500 s to reach Earth and 687 s to get to Mars. What was the distance from Mars to Earth?
   a. 5.6×10^7 km
   b. 5.6×10^{10} km
   c. 6.2×10^6 km
   d. 6.2×10^{12} km

9. Describe two ways in which light differs from all other

forms of wave energy.

a. 1. Light travels as a longitudinal wave.
   2. Light travels through a medium that fills up the empty space in the universe.

b. 1. Light travels as a transverse wave.
   2. Light travels through a medium that fills up the empty space in the universe.

c. 1. Light travels at the maximum possible speed in the universe.
   2. Light travels through a medium that fills up the empty space in the universe.

d. 1. Light travels at the maximum possible speed in the universe.
   2. Light does not require any material medium to travel.

10. Use the postulates of the special relativity theory to explain why the speed of light emitted from a fast-moving light source cannot exceed 3.00×10^{8} m/s.

a. The speed of light is maximum in the frame of reference of the moving object.
b. The speed of light is minimum in the frame of reference of the moving object.
c. The speed of light is the same in all frames of reference, including in the rest frame of its source.
d. Light always travels in a vacuum with a speed less than 3.00×10^{8} m/s, regardless of the speed of the source.

10.2 Consequences of Special Relativity

11. Halley's Comet comes near Earth every 75 years as it travels around its 22 billion km orbit at a speed of up to 700, 000 m/s. If it were possible to put a clock on the comet and read it each time the comet passed, which part of special relativity theory could be tested? What would be the expected result? Explain.

a. It would test time dilation. The clock would appear to be slightly slower.
b. It would test time dilation. The clock would appear to be slightly faster.
c. It would test length contraction. The length of the orbit would appear to be shortened from Earth's frame of reference.
d. It would test length contraction. The length of the orbit would appear to be shortened from the comet's frame of reference.

12. The nucleus of the isotope fluorine-18 (^{18}F) has mass defect of 2.44×10^{-28} kg. What is the binding energy of ^{18}F?

a. 2.2×10^{-11} J
b. 7.3×10^{-20} J
c. 2.2×10^{-20} J
d. 2.4×10^{-28} J

Problems

10.2 Consequences of Special Relativity

13. Deuterium (2 H) is an isotope of hydrogen that has one proton and one neutron in its nucleus. The binding energy of deuterium is 3.56×10^{-13} J. What is the mass defect of deuterium?

a. 3.20×10^{-4} kg
b. 1.68×10^{-6} kg
c. 1.19×10^{-21} kg
d. 3.96×10^{-30} kg

14. The sun orbits the center of the galaxy at a speed of 2.3×10^{5} m/s. The diameter of the sun is 1.391684×109 m. An observer is in a frame of reference that is stationary with respect to the center of the galaxy. True or false—The sun is moving fast enough for the observer to notice length contraction of the sun's diameter.

a. True
b. False

15. Consider the nuclear fission reaction $n + {}^{235}_{92}U \rightarrow {}^{144}_{56}Ba + {}^{89}_{36}Kr + 3n + E$. If a neutron has a rest mass of 1.009u, ${}^{235}_{92}U$ has a rest mass of 235.044u, ${}^{144}_{56}Ba$ has rest mass of 143.923u, and ${}^{89}_{36}Kr$ has a rest mass of 88.918u, what is the value of E in joules?

a. 1.8×10^{-11} J
b. 2.8×10^{-11} J
c. 1.8×10^{-10} J
d. 3.3×10^{-10} J

16. Consider the nuclear fusion reaction ${}^{2}_{1}H + {}^{3}_{1}H \rightarrow {}^{4}_{2}He + n + E$. If ${}^{2}_{1}H$ has a rest mass of 2.014u, ${}^{3}_{1}H$ has a rest mass of 3.016u, ${}^{4}_{2}He$ has a rest mass of 4.003u, and a neutron has a rest mass of 1.009u, what is the value of E in joules?

a. 2.7×10^{-14} J
b. 2.7×10^{-13} J
c. 2.7×10^{-12} J
d. 2.7×10^{-11} J

Performance Task

10.2 Consequences of Special Relativity

17. People are fascinated by the possibility of traveling across the universe to discover intelligent life on other planets. To do this, we would have to travel enormous distances. Suppose we could somehow travel at up to 90 percent of the speed of light. The closest star is Alpha Centauri, which is 4.37 light years away. (A light year is the distance light travels in one year.)

a. How long, from the point of view of people on Earth, would it take a space ship to travel to Alpha Centauri and back at 0.9*c*?
b. How much would the astronauts on the spaceship have aged by the time they got back to Earth?
c. Discuss the problems related to travel to stars that are 20 or 30 light years away. Assume travel speeds near the speed of light.

TEST PREP

Multiple Choice

10.1 Postulates of Special Relativity

18. What was the purpose of the Michelson–Morley experiment?
a. To determine the exact speed of light
b. To analyze the electromagnetic spectrum
c. To establish that Earth is the true frame of reference
d. To learn how the ether affected the propagation of light

19. What is the speed of light in a vacuum to three significant figures?
a. 1.86×10^5 m/s
b. 3.00×10^8 m/s
c. 6.71×10^8 m/s
d. 1.50×10^{11} m/s

20. How far does light travel in 1.00 min?
a. 1.80×10^7 km
b. 1.80×10^{13} km
c. 5.00×10^6 m
d. 5.00×10^8 m

21. Describe what is meant by the sentence, "Simultaneity is not absolute."
a. Events may appear simultaneous in all frames of reference.
b. Events may not appear simultaneous in all frames of reference.
c. The speed of light is not the same in all frames of reference.
d. The laws of physics may be different in different inertial frames of reference.

22. In 2003, Earth and Mars were aligned so that Earth was between Mars and the sun. Earth and Mars were 5.6×10^7 km from each other, which was the closest they had been in 50,000 years. People looking up saw Mars as a very bright red light on the horizon. If Mars was 2.06×10^8 km from the sun, how long did the reflected light people saw take to travel from the sun to Earth?
a. 14 min and 33 s
b. 12 min and 15 s
c. 11 min and 27 s
d. 3 min and 7 s

10.2 Consequences of Special Relativity

23. What does this expression represent: $\frac{1}{\sqrt{1-\frac{u^2}{c^2}}}$
a. time dilation
b. relativistic factor
c. relativistic energy
d. length contraction

24. What is the rest energy, E_0, of an object with a mass of 1.00 g ?
a. 3.00×10^5 J
b. 3.00×10^{11} J
c. 9.00×10^{13} J
d. 9.00×10^{16} J

25. The fuel rods in a nuclear reactor must be replaced from time to time because so much of the radioactive material has reacted that they can no longer produce energy. How would the mass of the spent fuel rods compare to their mass when they were new? Explain your answer.
a. The mass of the spent fuel rods would decrease.
b. The mass of the spent fuel rods would increase.
c. The mass of the spent fuel rods would remain the same.
d. The mass of the spent fuel rods would become close to zero.

Short Answer

10.1 Postulates of Special Relativity

26. What is the postulate having to do with the speed of light on which the theory of special relativity is based?
- a. The speed of light remains the same in all inertial frames of reference.
- b. The speed of light depends on the speed of the source emitting the light.
- c. The speed of light changes with change in medium through which it travels.
- d. The speed of light does not change with change in medium through which it travels.

27. What is the postulate having to do with reference frames on which the theory of special relativity is based?
- a. The frame of reference chosen is arbitrary as long as it is inertial.
- b. The frame of reference is chosen to have constant nonzero acceleration.
- c. The frame of reference is chosen in such a way that the object under observation is at rest.
- d. The frame of reference is chosen in such a way that the object under observation is moving with a constant speed.

28. If you look out the window of a moving car at houses going past, you sense that you are moving. What have you chosen as your frame of reference?
- a. the car
- b. the sun
- c. a house

29. Why did Michelson and Morley orient light beams at right angles to each other?
- a. To observe the particle nature of light
- b. To observe the effect of the passing ether on the speed of light
- c. To obtain a diffraction pattern by combination of light
- d. To obtain a constant path difference for interference of light

10.2 Consequences of Special Relativity

30. What is the relationship between the binding energy and the mass defect of an atomic nucleus?
- a. The binding energy is the energy equivalent of the mass defect, as given by $E_0 = mc$.
- b. The binding energy is the energy equivalent of the mass defect, as given by $E_0 = mc^2$.
- c. The binding energy is the energy equivalent of the mass defect, as given by $E_0 = \frac{m}{c}$
- d. The binding energy is the energy equivalent of the mass defect, as given by $E_0 = \frac{m}{c^2}$.

31. True or false—It is possible to just use the relationships $F = ma$ and $E = Fd$ to show that both sides of the equation $E_0 = mc^2$ have the same units.
- a. True
- b. False

32. Explain why the special theory of relativity caused the law of conservation of energy to be modified.
- a. The law of conservation of energy is not valid in relativistic mechanics.
- b. The law of conservation of energy has to be modified because of time dilation.
- c. The law of conservation of energy has to be modified because of length contraction.
- d. The law of conservation of energy has to be modified because of mass-energy equivalence.

33. The sun loses about 4×10^9 kg of mass every second. Explain in terms of special relativity why this is happening.
- a. The sun loses mass because of its high temperature.
- b. The sun loses mass because it is continuously releasing energy.
- c. The Sun loses mass because the diameter of the sun is contracted.
- d. The sun loses mass because the speed of the sun is very high and close to the speed of light.

Extended Response

10.1 Postulates of Special Relativity

34. Explain how Einstein's conclusion that nothing can travel faster than the speed of light contradicts an older concept about the speed of an object propelled from another, already moving, object.
- a. The older concept is that speeds are subtractive. For example, if a person throws a ball while running, the speed of the ball relative to the ground is the speed at which the person was running minus the speed of the throw. A relativistic example is when light is emitted from car headlights, it moves faster than the speed of light emitted from a stationary source.
- b. The older concept is that speeds are additive. For example, if a person throws a ball while running, the speed of the ball relative to the ground is the speed at which the person was running plus the speed of the throw. A relativistic example is when light is emitted from car headlights, it moves no

faster than the speed of light emitted from a stationary source. The car's speed does not affect the speed of light.

c. The older concept is that speeds are multiplicative. For example, if a person throws a ball while running, the speed of the ball relative to the ground is the speed at which the person was running multiplied by the speed of the throw. A relativistic example is when light is emitted from car headlights, it moves no faster than the speed of light emitted from a stationary source. The car's speed does not affect the speed of light.

d. The older concept is that speeds are frame independent. For example, if a person throws a ball while running, the speed of the ball relative to the ground has nothing to do with the speed at which the person was running. A relativistic example is when light is emitted from car headlights, it moves no faster than the speed of light emitted from a stationary source. The car's speed does not affect the speed of light.

35. A rowboat is drifting downstream. One person swims 20 m toward the shore and back, and another, leaving at the same time, swims upstream 20 m and back to the boat. The swimmer who swam toward the shore gets back first. Explain how this outcome is similar to the outcome expected in the Michelson–Morley experiment.

a. The rowboat represents Earth, the swimmers are beams of light, and the water is acting as the ether. Light going against the current of the ether would get back later because, by then, Earth would have moved on.

b. The rowboat represents the beam of light, the swimmers are the ether, and water is acting as Earth. Light going against the current of the ether would get back later because, by then, Earth would have moved on.

c. The rowboat represents the ether, the swimmers are ray of light, and the water is acting as the earth. Light going against the current of the ether would get back later because, by then, Earth would have moved on.

d. The rowboat represents the Earth, the swimmers are the ether, and the water is acting as the rays of light. Light going against the current of the ether would get back later because, by then, Earth would have moved on.

10.2 Consequences of Special Relativity

36. A helium-4 nucleus is made up of two neutrons and two protons. The binding energy of helium-4 is 4.53×10^{-12} J. What is the difference in the mass of this helium nucleus and the sum of the masses of two neutrons and two protons? Which weighs more, the nucleus or its constituents?

a. 1.51×10^{-20} kg; the constituents weigh more

b. 5.03×10^{-29} kg; the constituents weigh more

c. 1.51×10^{-29} kg; the nucleus weighs more

d. 5.03×10^{-29} kg; the nucleus weighs more

37. Use the equation for length contraction to explain the relationship between the length of an object perceived by a stationary observer who sees the object as moving, and the proper length of the object as measured in the frame of reference where it is at rest.

a. As the speed v of an object moving with respect to a stationary observer approaches c, the length perceived by the observer approaches zero. For other speeds, the length perceived is always less than the proper length.

b. As the speed v of an object moving with respect to a stationary observer approaches c, the length perceived by the observer approaches zero. For other speeds, the length perceived is always greater than the proper length.

c. As the speed v of an object moving with respect to a stationary observer approaches c, the length perceived by the observer approaches infinity. For other speeds, the length perceived is always less than the proper length.

d. As the speed v of an object moving with respect to a stationary observer approaches c, the length perceived by the observer approaches infinity. For other speeds, the length perceived is always greater than the proper length.

CHAPTER 11
Thermal Energy, Heat, and Work

Figure 11.1 The welder's gloves and helmet protect the welder from the electric arc, which transfers enough thermal energy to melt the rod, spray sparks, and emit high-energy electromagnetic radiation that can burn the retina of an unprotected eye. The thermal energy can be felt on exposed skin a few meters away, and its light can be seen for kilometers (Kevin S. O'Brien, U.S. Navy)

Chapter Outline

11.1 Temperature and Thermal Energy

11.2 Heat, Specific Heat, and Heat Transfer

11.3 Phase Change and Latent Heat

INTRODUCTION Heat is something familiar to all of us. We feel the warmth of the summer sun, the hot vapor rising up out of a cup of hot cocoa, and the cooling effect of our sweat. When we feel warmth, it means that heat is transferring energy *to* our bodies; when we feel cold, that means heat is transferring energy *away from* our bodies. Heat transfer is the movement of thermal energy from one place or material to another, and is caused by temperature differences. For example, much of our weather is caused by Earth evening out the temperature across the planet through wind and violent storms, which are driven by heat transferring energy away from the equator towards the cold poles. In this chapter, we'll explore the precise meaning of heat, how it relates to temperature as well as to other forms of energy, and its connection to work.

11.1 Temperature and Thermal Energy

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Explain that temperature is a measure of internal kinetic energy
- Interconvert temperatures between Celsius, Kelvin, and Fahrenheit scales

Section Key Terms

absolute zero	Celsius scale	degree Celsius ($^\circ\mathrm{C}$)	thermal energy
degree Fahrenheit ($^\circ\mathrm{F}$)	Fahrenheit scale	heat	
kelvin (K)	Kelvin scale	temperature	

Temperature

What is **temperature**? It's one of those concepts so ingrained in our everyday lives that, although we know what it means intuitively, it can be hard to define. It is tempting to say that temperature measures heat, but this is not strictly true. **Heat** is the transfer of energy due to a temperature difference. Temperature is defined in terms of the instrument we use to tell us how hot or cold an object is, based on a mechanism and scale invented by people. Temperature is literally defined as what we measure on a thermometer.

Heat is often confused with temperature. For example, we may say that the heat was unbearable, when we actually mean that the temperature was high. This is because we are sensitive to the flow of energy by heat, rather than the temperature. Since heat, like work, transfers energy, it has the SI unit of joule (J).

Atoms and molecules are constantly in motion, bouncing off one another in random directions. Recall that kinetic energy is the energy of motion, and that it increases in proportion to velocity squared. Without going into mathematical detail, we can say that **thermal energy**—the energy associated with heat—is the average kinetic energy of the particles (molecules or atoms) in a substance. Faster moving molecules have greater kinetic energies, and so the substance has greater thermal energy, and thus a higher temperature. The total internal energy of a system is the sum of the kinetic and potential energies of its atoms and molecules. Thermal energy is one of the subcategories of internal energy, as is chemical energy.

To measure temperature, some scale must be used as a standard of measurement. The three most commonly used temperature scales are the Fahrenheit, Celsius, and Kelvin scales. Both the **Fahrenheit scale** and **Celsius scale** are relative temperature scales, meaning that they are made around a reference point. For example, the Celsius scale uses the freezing point of water as its reference point; all measurements are either lower than the freezing point of water by a given number of degrees (and have a negative sign), or higher than the freezing point of water by a given number of degrees (and have a positive sign). The boiling point of water is 100 $^\circ\mathrm{C}$ for the Celsius scale, and its unit is the degree Celsius ($^\circ\mathrm{C}$).

On the Fahrenheit scale, the freezing point of water is at 32 $^\circ\mathrm{F}$, and the boiling point is at 212 $^\circ\mathrm{F}$. The unit of temperature on this scale is the degree Fahrenheit ($^\circ\mathrm{F}$). Note that the difference in degrees between the freezing and boiling points is greater for the Fahrenheit scale than for the Celsius scale. Therefore, a temperature difference of one degree Celsius is greater than a temperature difference of one degree Fahrenheit. Since 100 Celsius degrees span the same range as 180 Fahrenheit degrees, one degree on the Celsius scale is 1.8 times larger than one degree on the Fahrenheit scale (because $\frac{180}{100} = \frac{9}{5} = 1.8$). This relationship can be used to convert between temperatures in Fahrenheit and Celsius (see Figure 11.2).

Figure 11.2 Relationships between the Fahrenheit, Celsius, and Kelvin temperature scales, rounded to the nearest degree. The relative sizes of the scales are also shown.

The **Kelvin scale** is the temperature scale that is commonly used in science because it is an absolute temperature scale. This means that the theoretically lowest-possible temperature is assigned the value of zero. Zero degrees on the Kelvin scale is known as **absolute zero**; it is theoretically the point at which there is no molecular motion to produce thermal energy. On the original Kelvin scale first created by Lord Kelvin, all temperatures have positive values, making it useful for scientific work. The official temperature unit on this scale is the kelvin, which is abbreviated as K. The freezing point of water is 273.15 K, and the boiling point of water is 373.15 K.

Although absolute zero is possible in theory, it cannot be reached in practice. The lowest temperature ever created and measured during a laboratory experiment was 1.0×10^{-10} K, at Helsinki University of Technology in Finland. In comparison, the coldest recorded temperature for a place on Earth's surface was 183 K (−89 °C), at Vostok, Antarctica, and the coldest known place (outside the lab) in the universe is the Boomerang Nebula, with a temperature of 1 K. Luckily, most of us humans will never have to experience such extremes.

The average normal body temperature is 98.6 °F (37.0 °C), but people have been known to survive with body temperatures ranging from 75 °F to 111 °F (24 °C to 44 °C).

WATCH PHYSICS

Comparing Celsius and Fahrenheit Temperature Scales

This video shows how the Fahrenheit and Celsius temperature scales compare to one another.

Click to view content (https://www.openstax.org/l/02celfahtemp)

GRASP CHECK

Even without the number labels on the thermometer, you could tell which side is marked Fahrenheit and which is Celsius by how the degree marks are spaced. Why?

a. The separation between two consecutive divisions on the Fahrenheit scale is greater than a similar separation on the Celsius scale, because each degree Fahrenheit is equal to 1.8 degrees Celsius.
b. The separation between two consecutive divisions on the Fahrenheit scale is smaller than the similar separation on the Celsius scale, because each degree Celsius is equal to 1.8 degrees Fahrenheit.
c. The separation between two consecutive divisions on the Fahrenheit scale is greater than a similar separation on the Celsius scale, because each degree Fahrenheit is equal to 3.6 degrees Celsius.
d. The separation between two consecutive divisions on the Fahrenheit scale is smaller than a similar separation on the Celsius scale, because each degree Celsius is equal to 3.6 degrees Fahrenheit.

Converting Between Celsius, Kelvin, and Fahrenheit Scales

While the Fahrenheit scale is still the most commonly used scale in the United States, the majority of the world uses Celsius, and scientists prefer Kelvin. It's often necessary to convert between these scales. For instance, if the TV meteorologist gave the local weather report in kelvins, there would likely be some confused viewers! Table 11.1 gives the equations for conversion between the three temperature scales.

To Convert From...	Use This Equation
Celsius to Fahrenheit	$T_{°F} = \frac{9}{5}T_{°C} + 32$
Fahrenheit to Celsius	$T_{°C} = \frac{5}{9}(T_{°F} - 32)$
Celsius to Kelvin	$T_K = T_{°C} + 273.15$
Kelvin to Celsius	$T_{°C} = T_K - 273.15$
Fahrenheit to Kelvin	$T_K = \frac{5}{9}(T_{°F} - 32) + 273.15$
Kelvin to Fahrenheit	$T_{°F} = \frac{9}{5}(T_K - 273.15) + 32$

Table 11.1 Temperature Conversions

WORKED EXAMPLE

Room temperature is generally defined to be 25 °C. (a) What is room temperature in °F? (b) What is it in K?

STRATEGY

To answer these questions, all we need to do is choose the correct conversion equations and plug in the known values.

Solution for (a)

1. Choose the right equation. To convert from °C to °F, use the equation

$$T_{°F} = \frac{9}{5}T_{°C} + 32. \quad (11.1)$$

2. Plug the known value into the equation and solve.

$$T_{°F} = \frac{9}{5}25\ °C + 32 = 77\ °F \quad (11.2)$$

Solution for (b)

1. Choose the right equation. To convert from °C to K, use the equation

$$T_K = T_{°C} + 273.15. \quad (11.3)$$

2. Plug the known value into the equation and solve.

$$T_K = 25\ °C + 273.15 = 298K \quad (11.4)$$

Discussion

Living in the United States, you are likely to have more of a sense of what the temperature feels like if it's described as 77 °F than as 25 °C (or 298 K, for that matter).

WORKED EXAMPLE

Converting Between Temperature Scales: The Reaumur Scale

The Reaumur scale is a temperature scale that was used widely in Europe in the 18th and 19th centuries. On the Reaumur temperature scale, the freezing point of water is 0 °R and the boiling temperature is 80 °R. If "room temperature" is 25 °C on the Celsius scale, what is it on the Reaumur scale?

STRATEGY

To answer this question, we must compare the Reaumur scale to the Celsius scale. The difference between the freezing point and boiling point of water on the Reaumur scale is 80 °R . On the Celsius scale, it is 100 °C . Therefore, 100 °C = 80 °R . Both scales start at 0 ° for freezing, so we can create a simple formula to convert between temperatures on the two scales.

Solution

1. Derive a formula to convert from one scale to the other.

$$T_{°R} = \frac{0.80°R}{°C} \times T_{°C} \tag{11.5}$$

2. Plug the known value into the equation and solve.

$$T_{°R} = \frac{0.80°R}{°C} \times 25\ °C = 20°R \tag{11.6}$$

Discussion

As this example shows, relative temperature scales are somewhat arbitrary. If you wanted, you could create your own temperature scale!

Practice Problems

1. What is 12.0 °C in kelvins?
   a. 112.0 K
   b. 273.2 K
   c. 12.0 K
   d. 285.2 K

2. What is 32.0 °C in degrees Fahrenheit?
   a. 57.6 °F
   b. 25.6 °F
   c. 305.2 °F
   d. 89.6 °F

TIPS FOR SUCCESS

Sometimes it is not so easy to guess the temperature of the air accurately. Why is this? Factors such as humidity and wind speed affect how hot or cold we feel. Wind removes thermal energy from our bodies at a faster rate than usual, making us feel colder than we otherwise would; on a cold day, you may have heard the TV weather person refer to the *wind chill*. On humid summer days, people tend to feel hotter because sweat doesn't evaporate from the skin as efficiently as it does on dry days, when the evaporation of sweat cools us off.

Check Your Understanding

3. What is thermal energy?
   a. The thermal energy is the average potential energy of the particles in a system.
   b. The thermal energy is the total sum of the potential energies of the particles in a system.
   c. The thermal energy is the average kinetic energy of the particles due to the interaction among the particles in a system.
   d. The thermal energy is the average kinetic energy of the particles in a system.

4. What is used to measure temperature?

a. a galvanometer
b. a manometer
c. a thermometer
d. a voltmeter

11.2 Heat, Specific Heat, and Heat Transfer

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Explain heat, heat capacity, and specific heat
- Distinguish between conduction, convection, and radiation
- Solve problems involving specific heat and heat transfer

Section Key Terms

conduction	convection	heat capacity	radiation	specific heat

Heat Transfer, Specific Heat, and Heat Capacity

We learned in the previous section that temperature is proportional to the average kinetic energy of atoms and molecules in a substance, and that the average internal kinetic energy of a substance is higher when the substance's temperature is higher.

If two objects at different temperatures are brought in contact with each other, energy is transferred from the hotter object (that is, the object with the greater temperature) to the colder (lower temperature) object, until both objects are at the same temperature. There is no net heat transfer once the temperatures are equal because the amount of heat transferred from one object to the other is the same as the amount of heat returned. One of the major effects of heat transfer is temperature change: Heating increases the temperature while cooling decreases it. Experiments show that the heat transferred to or from a substance depends on three factors—the change in the substance's temperature, the mass of the substance, and certain physical properties related to the phase of the substance.

The equation for heat transfer Q is

$$Q = mc\Delta T, \tag{11.7}$$

where m is the mass of the substance and ΔT is the change in its temperature, in units of Celsius or Kelvin. The symbol c stands for **specific heat**, and depends on the material and phase. The specific heat is the amount of heat necessary to change the temperature of 1.00 kg of mass by 1.00 °C. The specific heat c is a property of the substance; its SI unit is J/(kg · K) or J/(kg · °C). The temperature change (ΔT) is the same in units of kelvins and degrees Celsius (but not degrees Fahrenheit). Specific heat is closely related to the concept of **heat capacity**. Heat capacity is the amount of heat necessary to change the temperature of a substance by 1.00 °C. In equation form, heat capacity C is $C = mc$, where m is mass and c is specific heat. Note that heat capacity is the same as specific heat, but without any dependence on mass. Consequently, two objects made up of the same material but with different masses will have different heat capacities. This is because the heat capacity is a property of an object, but specific heat is a property of *any* object made of the same material.

Values of specific heat must be looked up in tables, because there is no simple way to calculate them. Table 11.2 gives the values of specific heat for a few substances as a handy reference. We see from this table that the specific heat of water is five times that of glass, which means that it takes five times as much heat to raise the temperature of 1 kg of water than to raise the temperature of 1 kg of glass by the same number of degrees.

Substances	Specific Heat (c)
Solids	J/(kg · °C)
Aluminum	900

Table 11.2 Specific Heats of Various Substances.

Substances	Specific Heat (*c*)
Asbestos	800
Concrete, granite (average)	840
Copper	387
Glass	840
Gold	129
Human body (average)	3500
Ice (average)	2090
Iron, steel	452
Lead	128
Silver	235
Wood	1700
Liquids	
Benzene	1740
Ethanol	2450
Glycerin	2410
Mercury	139
Water	4186
Gases (at 1 atm constant pressure)	
Air (dry)	1015
Ammonia	2190
Carbon dioxide	833
Nitrogen	1040
Oxygen	913
Steam	2020

Table 11.2 Specific Heats of Various Substances.

Snap Lab

Temperature Change of Land and Water

What heats faster, land or water? You will answer this question by taking measurements to study differences in specific heat capacity.

- Open flame—Tie back all loose hair and clothing before igniting an open flame. Follow all of your teacher's instructions on how to ignite the flame. Never leave an open flame unattended. Know the location of fire safety equipment in the laboratory.

- Sand or soil
- Water
- Oven or heat lamp
- Two small jars
- Two thermometers

Instructions

Procedure

1. Place equal masses of dry sand (or soil) and water at the same temperature into two small jars. (The average density of soil or sand is about 1.6 times that of water, so you can get equal masses by using 50 percent more water by volume.)
2. Heat both substances (using an oven or a heat lamp) for the same amount of time.
3. Record the final temperatures of the two masses.
4. Now bring both jars to the same temperature by heating for a longer period of time.
5. Remove the jars from the heat source and measure their temperature every 5 minutes for about 30 minutes.

GRASP CHECK

Did it take longer to heat the water or the sand/soil to the same temperature? Which sample took longer to cool? What does this experiment tell us about how the specific heat of water compared to the specific heat of land?

a. The sand/soil will take longer to heat as well as to cool. This tells us that the specific heat of land is greater than that of water.
b. The sand/soil will take longer to heat as well as to cool. This tells us that the specific heat of water is greater than that of land.
c. The water will take longer to heat as well as to cool. This tells us that the specific heat of land is greater than that of water.
d. The water will take longer to heat as well as to cool. This tells us that the specific heat of water is greater than that of land.

Conduction, Convection, and Radiation

Whenever there is a temperature difference, heat transfer occurs. Heat transfer may happen rapidly, such as through a cooking pan, or slowly, such as through the walls of an insulated cooler.

There are three different heat transfer methods: **conduction**, **convection**, and **radiation**. At times, all three may happen simultaneously. See Figure 11.3.

Figure 11.3 In a fireplace, heat transfer occurs by all three methods: conduction, convection, and radiation. Radiation is responsible for most of the heat transferred into the room. Heat transfer also occurs through conduction into the room, but at a much slower rate. Heat transfer by convection also occurs through cold air entering the room around windows and hot air leaving the room by rising up the chimney.

Conduction is heat transfer through direct physical contact. Heat transferred between the electric burner of a stove and the bottom of a pan is transferred by conduction. Sometimes, we try to control the conduction of heat to make ourselves more comfortable. Since the rate of heat transfer is different for different materials, we choose fabrics, such as a thick wool sweater, that slow down the transfer of heat away from our bodies in winter.

As you walk barefoot across the living room carpet, your feet feel relatively comfortable…until you step onto the kitchen's tile floor. Since the carpet and tile floor are both at the same temperature, why does one feel colder than the other? This is explained by different rates of heat transfer: The tile material removes heat from your skin at a greater rate than the carpeting, which makes it *feel* colder.

Some materials simply conduct thermal energy faster than others. In general, metals (like copper, aluminum, gold, and silver) are good heat conductors, whereas materials like wood, plastic, and rubber are poor heat conductors.

Figure 11.4 shows particles (either atoms or molecules) in two bodies at different temperatures. The (average) kinetic energy of a particle in the hot body is higher than in the colder body. If two particles collide, energy transfers from the particle with greater kinetic energy to the particle with less kinetic energy. When two bodies are in contact, many particle collisions occur, resulting in a net flux of heat from the higher-temperature body to the lower-temperature body. The heat flux depends on the temperature difference $\Delta T = T_{\text{hot}} - T_{\text{cold}}$. Therefore, you will get a more severe burn from boiling water than from hot tap water.

Figure 11.4 The particles in two bodies at different temperatures have different average kinetic energies. Collisions occurring at the contact surface tend to transfer energy from high-temperature regions to low-temperature regions. In this illustration, a particle in the lower-temperature region (right side) has low kinetic energy before collision, but its kinetic energy increases after colliding with the contact

surface. In contrast, a particle in the higher-temperature region (left side) has more kinetic energy before collision, but its energy decreases after colliding with the contact surface.

Convection is heat transfer by the movement of a fluid. This type of heat transfer happens, for example, in a pot boiling on the stove, or in thunderstorms, where hot air rises up to the base of the clouds.

TIPS FOR SUCCESS

In everyday language, the term *fluid* is usually taken to mean liquid. For example, when you are sick and the doctor tells you to "push fluids," that only means to drink more beverages—not to breath more air. However, in physics, fluid means a liquid *or a gas*. Fluids move differently than solid material, and even have their own branch of physics, known as *fluid dynamics*, that studies how they move.

As the temperature of fluids increase, they expand and become less dense. For example, Figure 11.4 could represent the wall of a balloon with different temperature gases inside the balloon than outside in the environment. The hotter and thus faster moving gas particles inside the balloon strike the surface with more force than the cooler air outside, causing the balloon to expand. This decrease in density relative to its environment creates buoyancy (the tendency to rise). Convection is driven by buoyancy—hot air rises because it is less dense than the surrounding air.

Sometimes, we control the temperature of our homes or ourselves by controlling air movement. Sealing leaks around doors with weather stripping keeps out the cold wind in winter. The house in Figure 11.5 and the pot of water on the stove in Figure 11.6 are both examples of convection and buoyancy by human design. Ocean currents and large-scale atmospheric circulation transfer energy from one part of the globe to another, and are examples of natural convection.

Figure 11.5 Air heated by the so-called gravity furnace expands and rises, forming a convective loop that transfers energy to other parts of the room. As the air is cooled at the ceiling and outside walls, it contracts, eventually becoming denser than room air and sinking to the floor. A properly designed heating system like this one, which uses natural convection, can be quite efficient in uniformly heating a home.

Figure 11.6 Convection plays an important role in heat transfer inside this pot of water. Once conducted to the inside fluid, heat transfer to other parts of the pot is mostly by convection. The hotter water expands, decreases in density, and rises to transfer heat to other regions of the water, while colder water sinks to the bottom. This process repeats as long as there is water in the pot.

Radiation is a form of heat transfer that occurs when electromagnetic radiation is emitted or absorbed. Electromagnetic

radiation includes radio waves, microwaves, infrared radiation, visible light, ultraviolet radiation, X-rays, and gamma rays, all of which have different wavelengths and amounts of energy (shorter wavelengths have higher frequency and more energy).

You can feel the heat transfer from a fire and from the sun. Similarly, you can sometimes tell that the oven is hot without touching its door or looking inside—it may just warm you as you walk by. Another example is thermal radiation from the human body; people are constantly emitting infrared radiation, which is not visible to the human eye, but is felt as heat.

Radiation is the only method of heat transfer where no medium is required, meaning that the heat doesn't need to come into direct contact with or be transported by any matter. The space between Earth and the sun is largely empty, without any possibility of heat transfer by convection or conduction. Instead, heat is transferred by radiation, and Earth is warmed as it absorbs electromagnetic radiation emitted by the sun.

Figure 11.7 Most of the heat transfer from this fire to the observers is through infrared radiation. The visible light transfers relatively little thermal energy. Since skin is very sensitive to infrared radiation, you can sense the presence of a fire without looking at it directly. (Daniel X. O'Neil)

All objects absorb and emit electromagnetic radiation (see Figure 11.7). The rate of heat transfer by radiation depends mainly on the color of the object. Black is the most effective absorber and radiator, and white is the least effective. People living in hot climates generally avoid wearing black clothing, for instance. Similarly, black asphalt in a parking lot will be hotter than adjacent patches of grass on a summer day, because black absorbs better than green. The reverse is also true—black radiates better than green. On a clear summer night, the black asphalt will be colder than the green patch of grass, because black radiates energy faster than green. In contrast, white is a poor absorber and also a poor radiator. A white object reflects nearly all radiation, like a mirror.

Virtual Physics

Energy Forms and Changes

Click to view content (http://www.openstax.org/l/28energyForms)

In this animation, you will explore heat transfer with different materials. Experiment with heating and cooling the iron, brick, and water. This is done by dragging and dropping the object onto the pedestal and then holding the lever either to Heat or Cool. Drag a thermometer beside each object to measure its temperature—you can watch how quickly it heats or cools in real time.

Now let's try transferring heat between objects. Heat the brick and then place it in the cool water. Now heat the brick again, but then place it on top of the iron. What do you notice?

Selecting the fast forward option lets you speed up the heat transfers, to save time.

GRASP CHECK

Compare how quickly the different materials are heated or cooled. Based on these results, what material do you think has the greatest specific heat? Why? Which has the smallest specific heat? Can you think of a real-world situation where you would want to use an object with large specific heat?

a. Water will take the longest, and iron will take the shortest time to heat, as well as to cool. Objects with greater specific heat would be desirable for insulation. For instance, woolen clothes with large specific heat would prevent heat loss from the body.

b. Water will take the shortest, and iron will take the longest time to heat, as well as to cool. Objects with greater specific heat would be desirable for insulation. For instance, woolen clothes with large specific heat would prevent heat loss from the body.
c. Brick will take shortest and iron will take longest time to heat up as well as to cool down. Objects with greater specific heat would be desirable for insulation. For instance, woolen clothes with large specific heat would prevent heat loss from the body.
d. Water will take shortest and brick will take longest time to heat up as well as to cool down. Objects with greater specific heat would be desirable for insulation. For instance, woolen clothes with large specific heat would prevent heat loss from the body.

Solving Heat Transfer Problems

WORKED EXAMPLE

Calculating the Required Heat: Heating Water in an Aluminum Pan

A 0.500 kg aluminum pan on a stove is used to heat 0.250 L of water from 20.0 °C to 80.0 °C . (a) How much heat is required? What percentage of the heat is used to raise the temperature of (b) the pan and (c) the water?

STRATEGY

The pan and the water are always at the same temperature. When you put the pan on the stove, the temperature of the water and the pan is increased by the same amount. We use the equation for heat transfer for the given temperature change and masses of water and aluminum. The specific heat values for water and aluminum are given in the previous table.

Solution to (a)

Because the water is in thermal contact with the aluminum, the pan and the water are at the same temperature.

1. Calculate the temperature difference.

$$\Delta T = T_f - T_i = 60.0\ ^\circ\text{C} \tag{11.8}$$

2. Calculate the mass of water using the relationship between density, mass, and volume. Density is mass per unit volume, or $\rho = \frac{m}{V}$. Rearranging this equation, solve for the mass of water.

$$m_w = \rho \cdot V = 1000\ \text{kg/m}^3 \times \left(0.250\ \text{L} \times \frac{0.001\ \text{m}^3}{1\ \text{L}}\right) = 0.250\ \text{kg} \tag{11.9}$$

3. Calculate the heat transferred to the water. Use the specific heat of water in the previous table.

$$Q_w = m_w c_w \Delta T = (0.250\ \text{kg})(4186\ \text{J/kg°C})(60.0°\text{C}) = 62.8\ \text{kJ} \tag{11.10}$$

4. Calculate the heat transferred to the aluminum. Use the specific heat for aluminum in the previous table.

$$Q_{Al} = m_{Al} c_{Al} \Delta T = (0.500\ \text{kg})(900\ \text{J/kg°C})(60.0°\text{C}) = 27.0 \times 10^3\ \text{J} = 27.0\ \text{kJ} \tag{11.11}$$

5. Find the total transferred heat.

$$Q_{Total} = Q_w + Q_{Al} = 62.8\ \text{kJ} + 27.0\ \text{kJ} = 89.8\ \text{kJ} \tag{11.12}$$

Solution to (b)

The percentage of heat going into heating the pan is

$$\frac{27.0\ \text{kJ}}{89.8\ \text{kJ}} \times 100\% = 30.1\% \tag{11.13}$$

Solution to (c)

The percentage of heat going into heating the water is

$$\frac{62.8\ \text{kJ}}{89.8\ \text{kJ}} \times 100\% = 69.9\% \tag{11.14}$$

Discussion

In this example, most of the total heat transferred is used to heat the water, even though the pan has twice as much mass. This is

because the specific heat of water is over four times greater than the specific heat of aluminum. Therefore, it takes a bit more than twice as much heat to achieve the given temperature change for the water than for the aluminum pan.

Water can absorb a tremendous amount of energy with very little resulting temperature change. This property of water allows for life on Earth because it stabilizes temperatures. Other planets are less habitable because wild temperature swings make for a harsh environment. You may have noticed that climates closer to large bodies of water, such as oceans, are milder than climates landlocked in the middle of a large continent. This is due to the climate-moderating effect of water's large heat capacity—water stores large amounts of heat during hot weather and releases heat gradually when it's cold outside.

WORKED EXAMPLE

Calculating Temperature Increase: Truck Brakes Overheat on Downhill Runs

When a truck headed downhill brakes, the brakes must do work to convert the gravitational potential energy of the truck to internal energy of the brakes. This conversion prevents the gravitational potential energy from being converted into kinetic energy of the truck, and keeps the truck from speeding up and losing control. The increased internal energy of the brakes raises their temperature. When the hill is especially steep, the temperature increase may happen too quickly and cause the brakes to overheat.

Calculate the temperature increase of 100 kg of brake material with an average specific heat of 800 J/kg $\cdot$°C from a 10,000 kg truck descending 75.0 m (in vertical displacement) at a constant speed.

STRATEGY

We first calculate the gravitational potential energy (*Mgh*) of the truck, and then find the temperature increase produced in the brakes.

Solution

1. Calculate the change in gravitational potential energy as the truck goes downhill.

$$Mgh = (10,000\ \text{kg})\,(9.80\ \text{m/s}^2)\,(75.0\ \text{m}) = 7.35 \times 10^6\,\text{J} \tag{11.15}$$

2. Calculate the temperature change from the heat transferred by rearranging the equation $Q = mc\Delta T$ to solve for ΔT.

$$\Delta T = \frac{Q}{mc}, \tag{11.16}$$

where m is the mass of the brake material (not the entire truck). Insert the values $Q = 7.35\times10^6$ J (since the heat transfer is equal to the change in gravitational potential energy), $m = 100$ kg, and $c = 800$ J/kg $\cdot$ °C to find

$$\Delta T = \frac{7.35 \times 10^6\,\text{J}}{(100\ \text{kg})(800\ \text{J/kg}\ \cdot\ ^\circ\text{C})} = 91.9\ ^\circ\text{C}. \tag{11.17}$$

Discussion

This temperature is close to the boiling point of water. If the truck had been traveling for some time, then just before the descent, the brake temperature would likely be higher than the ambient temperature. The temperature increase in the descent would likely raise the temperature of the brake material above the boiling point of water, which would be hard on the brakes. This is why truck drivers sometimes use a different technique for called "engine braking" to avoid burning their brakes during steep descents. Engine braking is using the slowing forces of an engine in low gear rather than brakes to slow down.

Practice Problems

5. How much heat does it take to raise the temperature of 10.0 kg of water by 1.0 °C ?
   a. 84 J
   b. 42 J
   c. 84 kJ
   d. 42 kJ

6. Calculate the change in temperature of 1.0 kg of water that is initially at room temperature if 3.0 kJ of heat is added.
   a. 358 °C
   b. 716 °C
   c. 0.36 °C
   d. 0.72 °C

Check Your Understanding

7. What causes heat transfer?
   a. The mass difference between two objects causes heat transfer.
   b. The density difference between two objects causes heat transfer.
   c. The temperature difference between two systems causes heat transfer.
   d. The pressure difference between two objects causes heat transfer.

8. When two bodies of different temperatures are in contact, what is the overall direction of heat transfer?
   a. The overall direction of heat transfer is from the higher-temperature object to the lower-temperature object.
   b. The overall direction of heat transfer is from the lower-temperature object to the higher-temperature object.
   c. The direction of heat transfer is first from the lower-temperature object to the higher-temperature object, then back again to the lower-temperature object, and so-forth, until the objects are in thermal equilibrium.
   d. The direction of heat transfer is first from the higher-temperature object to the lower-temperature object, then back again to the higher-temperature object, and so-forth, until the objects are in thermal equilibrium.

9. What are the different methods of heat transfer?
   a. conduction, radiation, and reflection
   b. conduction, reflection, and convection
   c. convection, radiation, and reflection
   d. conduction, radiation, and convection

10. True or false—Conduction and convection cannot happen simultaneously
   a. True
   b. False

11.3 Phase Change and Latent Heat

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Explain changes in heat during changes of state, and describe latent heats of fusion and vaporization
- Solve problems involving thermal energy changes when heating and cooling substances with phase changes

Section Key Terms

condensation	freezing	latent heat	sublimation
latent heat of fusion	latent heat of vaporization	melting	vaporization
phase change	phase diagram	plasma	

Phase Changes

So far, we have learned that adding thermal energy by heat increases the temperature of a substance. But surprisingly, there are situations where adding energy does not change the temperature of a substance at all! Instead, the additional thermal energy acts to loosen bonds between molecules or atoms and causes a **phase change**. Because this energy enters or leaves a system during a phase change without causing a temperature change in the system, it is known as **latent heat** (latent means *hidden*).

The three phases of matter that you frequently encounter are solid, liquid and gas (see Figure 11.8). Solid has the least energetic state; atoms in solids are in close contact, with forces between them that allow the particles to vibrate but not change position with neighboring particles. (These forces can be thought of as springs that can be stretched or compressed, but not easily broken.)

Liquid has a more energetic state, in which particles can slide smoothly past one another and change neighbors, although they are still held together by their mutual attraction.

Gas has a more energetic state than liquid, in which particles are broken free of their bonds. Particles in gases are separated by distances that are large compared with the size of the particles.

The most energetic state of all is **plasma**. Although you may not have heard much about plasma, it is actually the most common state of matter in the universe—stars are made up of plasma, as is lightning. The plasma state is reached by heating a gas to the point where particles are pulled apart, separating the electrons from the rest of the particle. This produces an ionized gas that is a combination of the negatively charged free electrons and positively charged ions, known as plasma.

Figure 11.8 (a) Particles in a solid always have the same neighbors, held close by forces represented here by springs. These particles are essentially in contact with one another. A rock is an example of a solid. This rock retains its shape because of the forces holding its atoms or molecules together. (b) Particles in a liquid are also in close contact but can slide over one another. Forces between them strongly resist attempts to push them closer together and also hold them in close contact. Water is an example of a liquid. Water can flow, but it also remains in an open container because of the forces between its molecules. (c) Particles in a gas are separated by distances that are considerably larger than the size of the particles themselves, and they move about freely. A gas must be held in a closed container to prevent it from moving out into its surroundings. (d) The atmosphere is ionized in the extreme heat of a lightning strike.

During a phase change, matter changes from one phase to another, either through the addition of energy by heat and the transition to a more energetic state, or from the removal of energy by heat and the transition to a less energetic state.

Phase changes to a more energetic state include the following:

- **Melting**—Solid to liquid
- **Vaporization**—Liquid to gas (included boiling and evaporation)
- **Sublimation**—Solid to gas

Phase changes to a less energetic state are as follows:

- **Condensation**—Gas to liquid
- **Freezing**—Liquid to solid

Energy is required to melt a solid because the bonds between the particles in the solid must be broken. Since the energy involved in a phase changes is used to break bonds, there is no increase in the kinetic energies of the particles, and therefore no rise in temperature. Similarly, energy is needed to vaporize a liquid to overcome the attractive forces between particles in the liquid. There is no temperature change until a phase change is completed. The temperature of a cup of soda and ice that is initially at 0 °C stays at 0 °C until all of the ice has melted. In the reverse of these processes—freezing and condensation—energy is released

from the latent heat (see Figure 11.9).

Figure 11.9 (a) Energy is required to partially overcome the attractive forces between particles in a solid to form a liquid. That same energy must be removed for freezing to take place. (b) Particles are separated by large distances when changing from liquid to vapor, requiring significant energy to overcome molecular attraction. The same energy must be removed for condensation to take place. There is no temperature change until a phase change is completed. (c) Enough energy is added that the liquid state is skipped over completely as a substance undergoes sublimation.

The heat, Q, required to change the phase of a sample of mass m is

$Q = mL_f$ (for melting/freezing),

$Q = mL_v$ (for vaporization/condensation),

where L_f is the **latent heat of fusion**, and L_v is the **latent heat of vaporization**. The latent heat of fusion is the amount of heat needed to cause a phase change between solid and liquid. The latent heat of vaporization is the amount of heat needed to cause a

phase change between liquid and gas. L_f and L_v are coefficients that vary from substance to substance, depending on the strength of intermolecular forces, and both have standard units of J/kg. See Table 11.3 for values of L_f and L_v of different substances.

Substance	Melting Point (°C)	*Lf* (kJ/kg)	Boiling Point (°C)	*Lv* (kJ/kg)
Helium	−269.7	5.23	−268.9	20.9
Hydrogen	−259.3	58.6	−252.9	452
Nitrogen	−210.0	25.5	−195.8	201
Oxygen	−218.8	13.8	−183.0	213
Ethanol	−114	104	78.3	854
Ammonia	−78	332	−33.4	1370
Mercury	−38.9	11.8	357	272
Water	0.00	334	100.0	2256
Sulfur	119	38.1	444.6	326
Lead	327	24.5	1750	871
Antimony	631	165	1440	561
Aluminum	660	380	2520	11400
Silver	961	88.3	2193	2336
Gold	1063	64.5	2660	1578
Copper	1083	134	2595	5069
Uranium	1133	84	3900	1900
Tungsten	3410	184	5900	4810

Table 11.3 Latent Heats of Fusion and Vaporization, along with Melting and Boiling Points

Let's consider the example of adding heat to ice to examine its transitions through all three phases—solid to liquid to gas. A phase diagram indicating the temperature changes of water as energy is added is shown in Figure 11.10. The ice starts out at −20 °C, and its temperature rises linearly, absorbing heat at a constant rate until it reaches 0 °. Once at this temperature, the ice gradually melts, absorbing 334 kJ/kg. The temperature remains constant at 0 °C during this phase change. Once all the ice has melted, the temperature of the liquid water rises, absorbing heat at a new constant rate. At 100 °C, the water begins to boil and the temperature again remains constant while the water absorbs 2256 kJ/kg during this phase change. When all the liquid has become steam, the temperature rises again at a constant rate.

Figure 11.10 A graph of temperature versus added energy. The system is constructed so that no vapor forms while ice warms to become liquid water, and so when vaporization occurs, the vapor remains in the system. The long stretches of constant temperature values at 0 °C and 100 °C reflect the large latent heats of melting and vaporization, respectively.

We have seen that vaporization requires heat transfer to a substance from its surroundings. Condensation is the reverse process, where heat in transferred *away from* a substance *to* its surroundings. This release of latent heat increases the temperature of the surroundings. Energy must be removed from the condensing particles to make a vapor condense. This is why condensation occurs on cold surfaces: the heat transfers energy away from the warm vapor to the cold surface. The energy is exactly the same as that required to cause the phase change in the other direction, from liquid to vapor, and so it can be calculated from $Q = mL_v$. Latent heat is also released into the environment when a liquid freezes, and can be calculated from $Q = mL_f$.

FUN IN PHYSICS

Making Ice Cream

Figure 11.11 With the proper ingredients, some ice and a couple of plastic bags, you could make your own ice cream in five minutes. (ElinorD, Wikimedia Commons)

Ice cream is certainly easy enough to buy at the supermarket, but for the hardcore ice cream enthusiast, that may not be satisfying enough. Going through the process of making your own ice cream lets you invent your own flavors and marvel at the physics firsthand (Figure 11.11).

The first step to making homemade ice cream is to mix heavy cream, whole milk, sugar, and your flavor of choice; it could be as

simple as cocoa powder or vanilla extract, or as fancy as pomegranates or pistachios.

The next step is to pour the mixture into a container that is deep enough that you will be able to churn the mixture without it spilling over, and that is also freezer-safe. After placing it in the freezer, the ice cream has to be stirred vigorously every 45 minutes for four to five hours. This slows the freezing process and prevents the ice cream from turning into a solid block of ice. Most people prefer a soft creamy texture instead of one giant popsicle.

As it freezes, the cream undergoes a phase change from liquid to solid. By now, we're experienced enough to know that this means that the cream must experience a loss of heat. Where does that heat go? Due to the temperature difference between the freezer and the ice cream mixture, heat transfers thermal energy from the ice cream to the air in the freezer. Once the temperature in the freezer rises enough, the freezer is cooled by pumping excess heat outside into the kitchen.

A faster way to make ice cream is to chill it by placing the mixture in a plastic bag, surrounded by another plastic bag half full of ice. (You can also add a teaspoon of salt to the outer bag to lower the temperature of the ice/salt mixture.) Shaking the bag for five minutes churns the ice cream while cooling it evenly. In this case, the heat transfers energy out of the ice cream mixture and into the ice during the phase change.

This video (http://www.openstax.org/l/28icecream) gives a demonstration of how to make home-made ice cream using ice and plastic bags.

GRASP CHECK

Why does the ice bag method work so much faster than the freezer method for making ice cream?

a. Ice has a smaller specific heat than the surrounding air in a freezer. Hence, it absorbs more energy from the ice-cream mixture.
b. Ice has a smaller specific heat than the surrounding air in a freezer. Hence, it absorbs less energy from the ice-cream mixture.
c. Ice has a greater specific heat than the surrounding air in a freezer. Hence, it absorbs more energy from the ice-cream mixture.
d. Ice has a greater specific heat than the surrounding air in a freezer. Hence, it absorbs less energy from the ice-cream mixture.

Solving Thermal Energy Problems with Phase Changes

WORKED EXAMPLE

Calculating Heat Required for a Phase Change

Calculate a) how much energy is needed to melt 1.000 kg of ice at 0 °C (freezing point), and b) how much energy is required to vaporize 1.000 kg of water at 100 °C (boiling point).

STRATEGY FOR (A)

Using the equation for the heat required for melting, and the value of the latent heat of fusion of water from the previous table, we can solve for part (a).

Solution to (a)

The energy to melt 1.000 kg of ice is

$$Q = mL_f = (1.000\text{ kg})(334\text{ kJ/kg}) = 334\text{ kJ}. \tag{11.18}$$

STRATEGY FOR (B)

To solve part (b), we use the equation for heat required for vaporization, along with the latent heat of vaporization of water from the previous table.

Solution to (b)

The energy to vaporize 1.000 kg of liquid water is

$$Q = mL_v = (1.000\text{ kg})(2256\text{ kJ/kg}) = 2256\text{ kJ}. \tag{11.19}$$

Discussion

The amount of energy need to melt a kilogram of ice (334 kJ) is the same amount of energy needed to raise the temperature of 1.000 kg of liquid water from 0 °C to 79.8 °C . This example shows that the energy for a phase change is enormous compared to energy associated with temperature changes. It also demonstrates that the amount of energy needed for vaporization is even greater.

WORKED EXAMPLE

Calculating Final Temperature from Phase Change: Cooling Soda with Ice Cubes

Ice cubes are used to chill a soda at 20 °C and with a mass of $m_{soda} = 0.25 \text{ kg}$. The ice is at 0 °C and the total mass of the ice cubes is 0.018 kg. Assume that the soda is kept in a foam container so that heat loss can be ignored, and that the soda has the same specific heat as water. Find the final temperature when all of the ice has melted.

STRATEGY

The ice cubes are at the melting temperature of 0 °C . Heat is transferred from the soda to the ice for melting. Melting of ice occurs in two steps: first, the phase change occurs and solid (ice) transforms into liquid water at the melting temperature; then, the temperature of this water rises. Melting yields water at 0 °C, so more heat is transferred from the soda to this water until they are the same temperature. Since the amount of heat leaving the soda is the same as the amount of heat transferred to the ice.

$$Q_{ice} = -Q_{soda} \tag{11.20}$$

The heat transferred to the ice goes partly toward the phase change (melting), and partly toward raising the temperature after melting. Recall from the last section that the relationship between heat and temperature change is $Q = mc\Delta T$. For the ice, the temperature change is $T_f - 0 \text{ °C}$. The total heat transferred to the ice is therefore

$$Q_{ice} = m_{ice}L_f + m_{ice}c_w(T_f - 0 \text{ °C}). \tag{11.21}$$

Since the soda doesn't change phase, but only temperature, the heat given off by the soda is

$$Q_{soda} = m_{soda}c_w(T_f - 20 \text{ °C}). \tag{11.22}$$

Since $Q_{ice} = -Q_{soda}$,

$$m_{ice}L_f + m_{ice}c_w(T_f - 0 \text{ °C}) = -m_{soda}c_w(T_f - 20 \text{ °C}). \tag{11.23}$$

Bringing all terms involving T_f to the left-hand-side of the equation, and all other terms to the right-hand-side, we can solve for T_f .

$$T_f = \frac{m_{soda}c_w(20 \text{ °C}) - m_{ice}L_f}{(m_{soda} + m_{ice})c_w} \tag{11.24}$$

Substituting the known quantities

$$T_f = \frac{(0.25 \text{ kg})(4186 \text{ J/kg} \cdot \text{°C})(20 \text{ °C}) - (0.018 \text{ kg})(334{,}000 \text{ J/kg})}{(0.25 \text{ kg} + 0.018 \text{ kg})(4186 \text{ K/kg} \cdot \text{°C})} = 13 \text{ °C} \tag{11.25}$$

Discussion

This example shows the enormous energies involved during a phase change. The mass of the ice is about 7 percent the mass of the soda, yet it causes a noticeable change in the soda's temperature.

TIPS FOR SUCCESS

If the ice were not already at the freezing point, we would also have to factor in how much energy would go into raising its temperature up to 0 °C, before the phase change occurs. This would be a realistic scenario, because the temperature of ice is often below 0 °C .

Practice Problems

11. How much energy is needed to melt 2.00 kg of ice at 0 °C ?

a. 334 kJ
b. 336 kJ
c. 167 kJ
d. 668 kJ

12. If $2500\ \text{kJ}$ of energy is just enough to melt $3.0\ \text{kg}$ of a substance, what is the substance's latent heat of fusion?

a. $7500\ \text{kJ} \cdot \text{kg}$
b. $7500\ \text{kJ/kg}$
c. $830\ \text{kJ} \cdot \text{kg}$
d. $830\ \text{kJ/kg}$

Check Your Understanding

13. What is latent heat?

a. It is the heat that must transfer energy to or from a system in order to cause a mass change with a slight change in the temperature of the system.
b. It is the heat that must transfer energy to or from a system in order to cause a mass change without a temperature change in the system.
c. It is the heat that must transfer energy to or from a system in order to cause a phase change with a slight change in the temperature of the system.
d. It is the heat that must transfer energy to or from a system in order to cause a phase change without a temperature change in the system.

14. In which phases of matter are molecules capable of changing their positions?

a. gas, liquid, solid
b. liquid, plasma, solid
c. liquid, gas, plasma
d. plasma, gas, solid

KEY TERMS

absolute zero lowest possible temperature; the temperature at which all molecular motion ceases

Celsius scale temperature scale in which the freezing point of water is 0 °C and the boiling point of water is 100 °C at 1 atm of pressure

condensation phase change from gas to liquid

conduction heat transfer through stationary matter by physical contact

convection heat transfer by the movement of fluid

degree Celsius unit on the Celsius temperature scale

degree Fahrenheit unit on the Fahrenheit temperature scale

Fahrenheit scale temperature scale in which the freezing point of water is 32 °F and the boiling point of water is 212 °F

freezing phase change from liquid to solid

heat transfer of thermal (or internal) energy due to a temperature difference

heat capacity amount of heat necessary to change the temperature of a substance by 1.00 °C

Kelvin unit on the Kelvin temperature scale; note that it is never referred to in terms of "degrees" Kelvin

Kelvin scale temperature scale in which 0 K is the lowest possible temperature, representing absolute zero

latent heat heat related to the phase change of a substance rather than a change of temperature

latent heat of fusion amount of heat needed to cause a phase change between solid and liquid

latent heat of vaporization amount of heat needed to cause a phase change between liquid and gas

melting phase change from solid to liquid

phase change transition between solid, liquid, or gas states of a substance

plasma ionized gas that is a combination of the negatively charged free electrons and positively charged ions

radiation energy transferred by electromagnetic waves

specific heat amount of heat necessary to change the temperature of 1.00 kg of a substance by 1.00 °C

sublimation phase change from solid to gas

temperature quantity measured by a thermometer

thermal energy average random kinetic energy of a molecule or an atom

vaporization phase change from liquid to gas

SECTION SUMMARY

11.1 Temperature and Thermal Energy

- Temperature is the quantity measured by a thermometer.
- Temperature is related to the average kinetic energy of atoms and molecules in a system.
- Absolute zero is the temperature at which there is no molecular motion.
- There are three main temperature scales: Celsius, Fahrenheit, and Kelvin.
- Temperatures on one scale can be converted into temperatures on another scale.

11.2 Heat, Specific Heat, and Heat Transfer

- Heat is thermal (internal) energy transferred due to a temperature difference.
- The transfer of heat Q that leads to a change ΔT in the temperature of a body with mass m is $Q = mc\Delta T$, where c is the specific heat of the material.
- Heat is transferred by three different methods: conduction, convection, and radiation.
- Heat conduction is the transfer of heat between two objects in direct contact with each other.
- Convection is heat transfer by the movement of mass.
- Radiation is heat transfer by electromagnetic waves.

11.3 Phase Change and Latent Heat

- Most substances have four distinct phases: solid, liquid, gas, and plasma.
- Gas is the most energetic state and solid is the least.
- During a phase change, a substance undergoes transition to a higher energy state when heat is added, or to a lower energy state when heat is removed.
- Heat is added to a substance during melting and vaporization.
- Latent heat is released by a substance during condensation and freezing.
- Phase changes occur at fixed temperatures called boiling and freezing (or melting) points for a given substance.

KEY EQUATIONS

11.1 Temperature and Thermal Energy

Celsius to Fahrenheit conversion	$T_{°F} = \frac{9}{5}T_{°C} + 32$
Fahrenheit to Celsius conversion	$T_{°C} = \frac{5}{9}(T_{°F} - 32)$
Celsius to Kelvin conversion	$T_K = T_{°C} + 273.15$
Kelvin to Celsius conversion	$T_{°C} = T_K - 273.15$
Fahrenheit to Kelvin conversion	$T_K = \frac{5}{9}(T_{°F} - 32) + 273.15$
Kelvin to Fahrenheit conversion	$T_{°F} = \frac{9}{5}(T_K - 273.15) + 32$

11.2 Heat, Specific Heat, and Heat Transfer

heat transfer	$Q = mc\Delta T$
density	$\rho = \frac{m}{V}$

11.3 Phase Change and Latent Heat

heat transfer for melting/freezing phase change	$Q = mL_f$
heat transfer for vaporization/condensation phase change	$Q = mL_v$

CHAPTER REVIEW

Concept Items

11.1 Temperature and Thermal Energy

1. A glass of water has a temperature of 31 degrees Celsius. What state of matter is it in?
   a. solid
   b. liquid
   c. gas
   d. plasma

2. What is the difference between thermal energy and internal energy?
   a. The thermal energy of the system is the average kinetic energy of the system's constituent particles due to their motion. The total internal energy of the system is the sum of the kinetic energies and the potential energies of its constituent particles.
   b. The thermal energy of the system is the average potential energy of the system's constituent particles due to their motion. The total internal energy of the system is the sum of the kinetic energies and the potential energies of its constituent particles.
   c. The thermal energy of the system is the average kinetic energy of the system's constituent particles due to their motion. The total internal energy of the system is the sum of the kinetic energies of its constituent particles.
   d. The thermal energy of the system is the average potential energy of the systems' constituent particles due to their motion. The total internal energy of the system is the sum of the kinetic energies of its constituent particles.

3. What does the Celsius scale use as a reference point?
   a. The boiling point of mercury
   b. The boiling point of wax
   c. The freezing point of water
   d. The freezing point of wax

11.2 Heat, Specific Heat, and Heat Transfer

4. What are the SI units of specific heat?
   a. $J/kg^2 \cdot °C$
   b. $J \cdot kg^2/°C$
   c. $J \cdot kg/°C$
   d. $J/kg \cdot °C$

5. What is radiation?
   a. The transfer of energy through emission and absorption of the electromagnetic waves is known as radiation.
   b. The transfer of energy without any direct physical

contact between any two substances.
c. The transfer of energy through direct physical contact between any two substances.
d. The transfer of energy by means of the motion of fluids at different temperatures and with different densities.

11.3 Phase Change and Latent Heat

6. Why is there no change in temperature during a phase change, even if energy is absorbed by the system?
a. The energy is used to break bonds between particles, and so does not increase the potential energy of the system's particles.
b. The energy is used to break bonds between particles, and so increases the potential energy of the system's particles.
c. The energy is used to break bonds between particles, and so does not increase the kinetic energy of the system's particles.
d. The energy is used to break bonds between particles, and so increases the kinetic energy of the system's particles.

7. In which two phases of matter do atoms and molecules have the most distance between them?
a. gas and solid
b. gas and liquid
c. gas and plasma
d. liquid and plasma

Critical Thinking Items

11.1 Temperature and Thermal Energy

8. The temperature of two equal quantities of water needs to be raised - the first container by 5 degrees Celsius and the second by 5 degrees Fahrenheit. Which one would require more heat?
a. The heat required by the first container is more than the second because each degree Celsius is equal to 1.8 degrees Fahrenheit.
b. The heat required by the first container is less than the second because each degree Fahrenheit is equal to 1.8 degrees Celsius.
c. The heat required by the first container is more than the second because each degree Celsius is equal to 3.6 degrees Fahrenheit.
d. The heat required by the first container is less than the second because each degree Fahrenheit is equal to 3.6 degrees Celsius.

9. What is 100.00 °C in kelvins?
a. 212.00 K
b. 100.00 K
c. 473.15 K
d. 373.15 K

11.2 Heat, Specific Heat, and Heat Transfer

10. The value of specific heat is the same whether the units are J/kg·K or J/kg·°C. How?
a. Temperature difference is dependent on the chosen temperature scale.
b. Temperature change is different in units of kelvins and degrees Celsius.
c. Reading of temperatures in kelvins and degree Celsius are the same.
d. The temperature change is the same in units of kelvins and degrees Celsius.

11. If the thermal energy of a perfectly black object is increased by conduction, will the object remain black in appearance? Why or why not?
a. No, the energy of the radiation increases as the temperature increases, and the radiation becomes visible at certain temperatures.
b. Yes, the energy of the radiation decreases as the temperature increases, and the radiation remains invisible at those energies.
c. No, the energy of the radiation decreases as the temperature increases, until the frequencies of the radiation are the same as those of visible light.
d. Yes, as the temperature increases, and the energy is transferred from the object by other mechanisms besides radiation, so that the energy of the radiation does not increase.

12. What is the specific heat of a substance that requires 5.00 kJ of heat to raise the temperature of 3.00 kg by 5.00 °F?
a. 3.33×103 J/kg ·° C
b. 6.00×103 J/kg ·° C
c. 3.33×102 J/kg · ° C
d. 6.00×102 J/kg · ° C

11.3 Phase Change and Latent Heat

13. Assume 1.0 kg of ice at 0 °C starts to melt. It absorbs 300 kJ of energy by heat. What is the temperature of the water afterwards?
a. 10 °C
b. 20 °C
c. 5 °C
d. 0 °C

Problems

11.1 Temperature and Thermal Energy

14. What is 35.0 °F in kelvins?
   a. 1.67 K
   b. 35.0 K
   c. -271.5 K
   d. 274.8 K

15. Design a temperature scale where the freezing point of water is 0 degrees and its boiling point is 70 degrees. What would be the room temperature on this scale?
   a. If room temperature is 25.0 °C, the temperature on the new scale will be 17.5 °.
   b. If room temperature is 25.0 °C, the temperature on the new scale will be 25.0°.
   c. If the room temperature is 25.0 °C, the temperature on the new scale will be 35.7°.
   d. If the room temperature is 25.0 °C, the temperature on the new scale will be 50.0°.

11.2 Heat, Specific Heat, and Heat Transfer

16. A certain quantity of water is given 4.0 kJ of heat. This raises its temperature by 30.0 °F. What is the mass of the water in grams?
   a. 5.7 g
   b. 570 g
   c. 5700 g
   d. 57 g

17. 5290 J of heat is given to 0.500 kg water at 15.00 °C. What will its final temperature be?
   a. 15.25° C
   b. 12.47 ° C
   c. 40.3° C
   d. 17.53° C

11.3 Phase Change and Latent Heat

18. How much energy would it take to heat 1.00 kg of ice at 0 °C to water at 15.0 °C?
   a. 271 kJ
   b. 334 kJ
   c. 62.8 kJ
   d. 397 kJ

19. Ice cubes are used to chill a soda with a mass m_{soda} = 0.300 kg at 15.0 °C. The ice is at 0 °C, and the total mass of the ice cubes is 0.020 kg. Assume that the soda is kept in a foam container so that heat loss can be ignored, and that the soda has the same specific heat as water. Find the final temperature when all ice has melted.
   a. 19.02 °C
   b. 90.3 °C
   c. 0.11 °C
   d. 9.03 °C

Performance Task

11.3 Phase Change and Latent Heat

20. You have been tasked with designing a baking pan that will bake batter the fastest. There are four materials available for you to test.
   - Four pans of similar design, consisting of aluminum, iron (steel), copper, and glass
   - Oven or similar heating source
   - Device for measuring high temperatures
   - Balance for measuring mass

   Instructions

   Procedure

   1. Design a safe experiment to test the specific heat of each material (i.e., no extreme temperatures should be used)
   2. Write down the materials needed for your experiment and the procedure you will follow. Make sure that you include every detail, so that the experiment can be repeated by others.
   3. Carry out the experiment and record any data collected.
   4. Review your results and make a recommendation as to which metal should be used for the pan.
      a. What physical quantities do you need to measure to determine the specific heats for the different materials?
      b. How does the glass differ from the metals in terms of thermal properties?
      c. What are your sources of error?

TEST PREP

Multiple Choice

11.1 Temperature and Thermal Energy

21. The temperature difference of 1 K is the same as
   a. 1 degree Celsius
   b. 1 degree Fahrenheit
   c. 273.15 degrees Celsius
   d. 273.15 degrees Fahrenheit

22. What is the preferred temperature scale used in scientific laboratories?
a. celsius
b. fahrenheit
c. kelvin
d. rankine

11.2 Heat, Specific Heat, and Heat Transfer

23. Which phase of water has the largest specific heat?
a. solid
b. liquid
c. gas

24. What kind of heat transfer requires no medium?
a. conduction
b. convection
c. reflection
d. radiation

25. Which of these substances has the greatest specific heat?
a. copper
b. mercury
c. aluminum
d. wood

26. Give an example of heat transfer through convection.
a. The energy emitted from the filament of an electric bulb
b. The energy coming from the sun
c. A pan on a hot burner
d. Water boiling in a pot

11.3 Phase Change and Latent Heat

27. What are the SI units of latent heat?
a. J/kg
b. J.kg
c. J/cal
d. cal/kg

28. Which substance has the largest latent heat of fusion?
a. gold
b. water
c. mercury
d. tungsten

29. In which phase changes does matter undergo a transition to a more energetic state?
a. freezing and vaporization
b. melting and sublimation
c. melting and vaporization
d. melting and freezing

30. A room has a window made from thin glass. The room is colder than the air outside. There is some condensation on the glass window. On which side of the glass would the condensation most likely be found?
a. Condensation is on the outside of the glass when the cool, dry air outside the room comes in contact with the cold pane of glass.
b. Condensation is on the outside of the glass when the warm, moist air outside the room comes in contact with the cold pane of glass.
c. Condensation is on the inside of the glass when the cool, dry air inside the room comes in contact with the cold pane of glass.
d. Condensation is on the inside of the glass when the warm, moist air inside the room comes in contact with the cold pane of glass.

Short Answer

11.1 Temperature and Thermal Energy

31. What is *absolute zero* on the Fahrenheit scale?
a. 0 °F
b. 32 °F
c. -273.15 °F
d. -459.67 °F

32. What is *absolute zero* on the Celsius scale?
a. 0 °C
b. 273.15 °C
c. -459.67 °C
d. -273.15 °C

33. A planet's atmospheric pressure is such that water there boils at a lower temperature than it does at sea level on Earth. If a Celsius scale is derived on this planet, will it be the same as that on Earth?
a. The Celsius scale derived on the planet will be the same as that on Earth, because the Celsius scale is independent of the freezing and boiling points of water.
b. The Celsius scale derived on that planet will not be the same as that on Earth, because the Celsius scale is dependent and derived by using the freezing and boiling points of water.
c. The Celsius scale derived on the planet will be the same as that on Earth, because the Celsius scale is an absolute temperature scale based on molecular motion, which is independent of pressure.
d. The Celsius scale derived on the planet will not be the same as that on Earth, but the Fahrenheit scale

will be the same, because its reference temperatures are not based on the freezing and boiling points of water.

34. What is the difference between the freezing point and boiling point of water on the Reaumur scale?
a. The boiling point of water is 80° on the Reaumur scale.
b. Reaumur scale is less than 120°.
c. 100°
d. 80°

11.2 Heat, Specific Heat, and Heat Transfer

35. In the specific heat equation what does *c* stand for?
a. Total heat
b. Specific heat
c. Specific temperature
d. Specific mass

36. Specific heat may be measured in J/kg · K, J/kg · °C. What other units can it be measured in?
a. kg/kcal · °C
b. kcal · °C/kg
c. kg · °C/kcal
d. kcal/kg · °C

37. What is buoyancy?
a. Buoyancy is a downward force exerted by a solid that opposes the weight of an object.
b. Buoyancy is a downward force exerted by a fluid that opposes the weight of an immersed object.
c. Buoyancy is an upward force exerted by a solid that opposes the weight an object.
d. Buoyancy is an upward force exerted by a fluid that opposes the weight of an immersed object.

38. Give an example of convection found in nature.
a. heat transfer through metallic rod
b. heat transfer from the sun to Earth
c. heat transfer through ocean currents
d. heat emitted by a light bulb into its environment

39. Calculate the temperature change in a substance with specific heat 735 J/kg · °C when 14 kJ of heat is given to a 3.0-kg sample of that substance.
a. 57 °C
b. 63 °C
c. 1.8×10^{-2} °C
d. 6.3 °C

40. Aluminum has a specific heat of 900 J/kg·°C. How much energy would it take to change the temperature of 2 kg aluminum by 3 °C?
a. 1.3 kJ
b. 0.60 kJ
c. 54 kJ
d. 5.4 kJ

11.3 Phase Change and Latent Heat

41. Upon what does the required amount of heat removed to freeze a sample of a substance depend?
a. The mass of the substance and its latent heat of vaporization
b. The mass of the substance and its latent heat of fusion
c. The mass of the substance and its latent heat of sublimation
d. The mass of the substance only

42. What do latent heats, L_f and L_v, depend on?
a. L_f and L_v depend on the forces between the particles in the substance.
b. L_f and L_v depend on the mass of the substance.
c. L_f and L_v depend on the volume of the substance.
d. L_f and L_v depend on the temperature of the substance.

43. How much energy is required to melt 7.00 kg a block of aluminum that is at its melting point? (Latent heat of fusion of aluminum is 380 kJ/kg.)
a. 54.3 kJ
b. 2.66 kJ
c. 0.0184 kJ
d. 2.66×10^{3} kJ

44. A 3.00 kg sample of a substance is at its boiling point. If 5,360 kJ of energy are enough to boil away the entire substance, what is its latent heat of vaporization?
a. 2,685 kJ/kg
b. 3,580 kJ/kg
c. 895 kJ/kg
d. 1,790 kJ/kg

Extended Response

11.1 Temperature and Thermal Energy

45. What is the meaning of absolute zero?
a. It is the temperature at which the internal energy of the system is maximum, because the speed of its constituent particles increases to maximum at this point.
b. It is the temperature at which the internal energy of the system is maximum, because the speed of its constituent particles decreases to zero at this point.

c. It is the temperature at which the internal energy of the system approaches zero, because the speed of its constituent particles increases to a maximum at this point.
d. It is that temperature at which the internal energy of the system approaches zero, because the speed of its constituent particles decreases to zero at this point.

46. Why does it feel hotter on more humid days, even though there is no difference in temperature?
a. On hot, dry days, the evaporation of the sweat from the skin cools the body, whereas on humid days the concentration of water in the atmosphere is lower, which reduces the evaporation rate from the skin's surface.
b. On hot, dry days, the evaporation of the sweat from the skin cools the body, whereas on humid days the concentration of water in the atmosphere is higher, which reduces the evaporation rate from the skin's surface.
c. On hot, dry days, the evaporation of the sweat from the skin cools the body, whereas on humid days the concentration of water in the atmosphere is lower, which increases the evaporation rate from the skin's surface.
d. On hot, dry days, the evaporation of the sweat from the skin cools the body, whereas on humid days the concentration of water in the atmosphere is higher, which increases the evaporation rate from the skin's surface.

11.2 Heat, Specific Heat, and Heat Transfer

47. A hot piece of metal needs to be cooled. If you were to put the metal in ice or in cold water, such that the ice did not melt and the temperature of either changed by the same amount, which would reduce the metal's temperature more? Why?
a. Water would reduce the metal's temperature more, because water has a greater specific heat than ice.
b. Water would reduce the metal's temperature more, because water has a smaller specific heat than ice.
c. Ice would reduce the metal's temperature more, because ice has a smaller specific heat than water.
d. Ice would reduce the metal's temperature more, because ice has a greater specific heat than water.

48. On a summer night, why does a black object seem colder than a white one?
a. The black object radiates energy faster than the white one, and hence reaches a lower temperature in less time.
b. The black object radiates energy slower than the white one, and hence reaches a lower temperature in less time.
c. The black object absorbs energy faster than the white one, and hence reaches a lower temperature in less time.
d. The black object absorbs energy slower than the white one, and hence reaches a lower temperature in less time.

49. Calculate the difference in heat required to raise the temperatures of 1.00 kg of gold and 1.00 kg of aluminum by 1.00 °C. (The specific heat of aluminum equals 900 J/kg · °C; the specific heat of gold equals 129 J/kg · °C.)
a. 771 J
b. 129 J
c. 90 J
d. 900 J

11.3 Phase Change and Latent Heat

50. True or false—You have an ice cube floating in a glass of water with a thin thread resting across the cube. If you cover the ice cube and thread with a layer of salt, they will stick together, so that you are able to lift the ice-cube when you pick up the thread.
a. True
b. False

51. Suppose the energy required to freeze 0.250 kg of water were added to the same mass of water at an initial temperature of 1.0 °C. What would be the final temperature of the water?
a. -69.8 °C
b. 79.8 °C
c. -78.8 °C
d. 80.8 °C

CHAPTER 12
Thermodynamics

Figure 12.1 A steam engine uses energy transfer by heat to do work. (Modification of work by Gerald Friedrich, Pixabay)

Chapter Outline

12.1 Zeroth Law of Thermodynamics: Thermal Equilibrium

12.2 First law of Thermodynamics: Thermal Energy and Work

12.3 Second Law of Thermodynamics: Entropy

12.4 Applications of Thermodynamics: Heat Engines, Heat Pumps, and Refrigerators

INTRODUCTION Energy can be transferred to or from a system, either through a temperature difference between it and another system (i.e., by heat) or by exerting a force through a distance (work). In these ways, energy can be converted into other forms of energy in other systems. For example, a car engine burns fuel for heat transfer into a gas. Work is done by the gas as it exerts a force through a distance by pushing a piston outward. This work converts the energy into a variety of other forms—into an increase in the car's kinetic or gravitational potential energy; into electrical energy to run the spark plugs, radio, and lights; and back into stored energy in the car's battery. But most of the thermal energy transferred by heat from the fuel burning in the engine does not do work on the gas. Instead, much of this energy is released into the surroundings at lower temperature (i.e., lost through heat), which is quite inefficient. Car engines are only about 25 to 30 percent efficient. This inefficiency leads to increased fuel costs, so there is great interest in improving fuel efficiency. However, it is common knowledge that modern gasoline engines cannot be made much more efficient. The same is true about the conversion to electrical energy in large power stations, whether they are coal, oil, natural gas, or nuclear powered. Why is this the case?

The answer lies in the nature of heat. Basic physical laws govern how heat transfer for doing work takes place and limit the

maximum possible efficiency of the process. This chapter will explore these laws as well their applications to everyday machines. These topics are part of *thermodynamics*—the study of heat and its relationship to doing work.

12.1 Zeroth Law of Thermodynamics: Thermal Equilibrium

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Explain the zeroth law of thermodynamics

Section Key Terms

thermal equilibrium	zeroth law of thermodynamics

We learned in the previous chapter that when two objects (or *systems*) are in contact with one another, heat will transfer thermal energy from the object at higher temperature to the one at lower temperature until they both reach the same temperature. The objects are then in **thermal equilibrium**, and no further temperature changes will occur if they are isolated from other systems. The systems interact and change because their temperatures are different, and the changes stop once their temperatures are the same. Thermal equilibrium is established when two bodies are in *thermal contact* with each other—meaning heat transfer (i.e., the transfer of energy by heat) can occur between them. If two systems cannot freely exchange energy, they will not reach thermal equilibrium. (It is fortunate that empty space stands between Earth and the sun, because a state of thermal equilibrium with the sun would be too toasty for life on this planet!)

If two systems, A and B, are in thermal equilibrium with each another, and B is in thermal equilibrium with a third system, C, then A is also in thermal equilibrium with C. This statement may seem obvious, because all three have the same temperature, but it is basic to thermodynamics. It is called the **zeroth law of thermodynamics**.

TIPS FOR SUCCESS

The zeroth law of thermodynamics is very similar to the transitive property of equality in mathematics: If a = b and b = c, then a = c.

You may be wondering at this point, why the wacky name? Shouldn't this be called the *first* law of thermodynamics rather than the *zeroth*? The explanation is that this law was discovered after the first and second laws of thermodynamics but is so fundamental that scientists decided it should logically come first.

As an example of the zeroth law in action, consider newborn babies in neonatal intensive-care units in hospitals. Prematurely born or sick newborns are placed in special incubators. These babies have very little covering while in the incubators, so to an observer, they look as though they may not be warm enough. However, inside the incubator, the temperature of the air, the cot, and the baby are all the same—that is, they are in thermal equilibrium. The ambient temperature is just high enough to keep the baby safe and comfortable.

WORK IN PHYSICS

Thermodynamics Engineer

Thermodynamics engineers apply the principles of thermodynamics to mechanical systems so as to create or test products that rely on the interactions between heat, work, pressure, temperature, and volume. This type of work typically takes place in the aerospace industry, chemical manufacturing companies, industrial manufacturing plants, power plants (Figure 12.2), engine manufacturers, or electronics companies.

Figure 12.2 An engineer makes a site visit to the Baghdad South power plant.

The need for energy creates quite a bit of demand for thermodynamics engineers, because both traditional energy companies and alternative (*green*) energy startups rely on interactions between heat and work and so require the expertise of thermodynamics engineers. Traditional energy companies use mainly nuclear energy and energy from burning fossil fuels, such as coal. Alternative energy is finding new ways to harness renewable and, often, more readily available energy sources, such as solar, water, wind, and bio-energy.

A thermodynamics engineer in the energy industry can find the most efficient way to turn the burning of a biofuel or fossil fuel into energy, store that energy for times when it's needed most, or figure out how to best deliver that energy from where it's produced to where it's used: in homes, factories, and businesses. Additionally, he or she might also design pollution-control equipment to remove harmful pollutants from the smoke produced as a by-product of burning fuel. For example, a thermodynamics engineer may develop a way to remove mercury from burning coal in a coal-fired power plant.

Thermodynamics engineering is an expanding field, where employment opportunities are expected to grow by as much as 27 percent between 2012 and 2022, according to the U.S. Bureau of Labor Statistics. To become a thermodynamics engineer, you must have a college degree in chemical engineering, mechanical engineering, environmental engineering, aerospace engineering, civil engineering, or biological engineering (depending on which type of career you wish to pursue), with coursework in physics and physical chemistry that focuses on thermodynamics.

GRASP CHECK

What would be an example of something a thermodynamics engineer would do in the aeronautics industry?

a. Test the fuel efficiency of a jet engine
b. Test the functioning of landing gear
c. Test the functioning of a lift control device
d. Test the autopilot functions

Check Your Understanding

1. What is thermal equilibrium?
   a. When two objects in contact with each other are at the same pressure, they are said to be in thermal equilibrium.
   b. When two objects in contact with each other are at different temperatures, they are said to be in thermal equilibrium.
   c. When two objects in contact with each other are at the same temperature, they are said to be in thermal equilibrium.
   d. When two objects not in contact with each other are at the same pressure, they are said to be in thermal equilibrium.

2. What is the zeroth law of thermodynamics?

a. Energy can neither be created nor destroyed in a chemical reaction.
b. If two systems, A and B, are in thermal equilibrium with each another, and B is in thermal equilibrium with a third system, C, then A is also in thermal equilibrium with C.
c. Entropy of any isolated system not in thermal equilibrium always increases.
d. Entropy of a system approaches a constant value as temperature approaches absolute zero.

12.2 First law of Thermodynamics: Thermal Energy and Work

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Describe how pressure, volume, and temperature relate to one another and to work, based on the ideal gas law
- Describe pressure–volume work
- Describe the first law of thermodynamics verbally and mathematically
- Solve problems involving the first law of thermodynamics

Section Key Terms

Boltzmann constant	first law of thermodynamics	ideal gas law	internal energy	pressure

Pressure, Volume, Temperature, and the Ideal Gas Law

Before covering the first law of thermodynamics, it is first important to understand the relationship between **pressure**, volume, and temperature. Pressure, P, is defined as

$$P = \frac{F}{A}, \tag{12.1}$$

where F is a force applied to an area, A, that is perpendicular to the force.

Depending on the area over which it is exerted, a given force can have a significantly different effect, as shown in Figure 12.3.

Figure 12.3 (a) Although the person being poked with the finger might be irritated, the force has little lasting effect. (b) In contrast, the same force applied to an area the size of the sharp end of a needle is great enough to break the skin.

The SI unit for pressure is the *pascal*, where $1\text{ Pa} = 1\text{ N/m}^2$.

Pressure is defined for all states of matter but is particularly important when discussing fluids (such as air). You have probably heard the word *pressure* being used in relation to blood (high or low blood pressure) and in relation to the weather (high- and low-pressure weather systems). These are only two of many examples of pressures in fluids.

The relationship between the pressure, volume, and temperature for an ideal gas is given by the **ideal gas law**. A gas is considered ideal at low pressure and fairly high temperature, and forces between its component particles can be ignored. The ideal gas law states that

$$PV = NkT. \tag{12.2}$$

where P is the pressure of a gas, V is the volume it occupies, N is the number of particles (atoms or molecules) in the gas, and T is

its absolute temperature. The constant k is called the **Boltzmann constant** and has the value $k = 1.38 \times 10^{-23}$ J/K, For the purposes of this chapter, we will not go into calculations using the ideal gas law. Instead, it is important for us to notice from the equation that the following are true for a given mass of gas:

- When volume is constant, pressure is directly proportional to temperature.
- When temperature is constant, pressure is inversely proportional to volume.
- When pressure is constant, volume is directly proportional to temperature.

This last point describes *thermal expansion*—the change in size or volume of a given mass with temperature. What is the underlying cause of thermal expansion? An increase in temperature means that there's an increase in the kinetic energy of the individual atoms. Gases are especially affected by thermal expansion, although liquids expand to a lesser extent with similar increases in temperature, and even solids have minor expansions at higher temperatures. This is why railroad tracks and bridges have expansion joints that allow them to freely expand and contract with temperature changes.

To get some idea of how pressure, temperature, and volume of a gas are related to one another, consider what happens when you pump air into a deflated tire. The tire's volume first increases in direct proportion to the amount of air injected, without much increase in the tire pressure. Once the tire has expanded to nearly its full size, the walls limit volume expansion. If you continue to pump air into tire (which now has a nearly constant volume), the pressure increases with increasing temperature (see Figure 12.4).

Figure 12.4 (a) When air is pumped into a deflated tire, its volume first increases without much increase in pressure. (b) When the tire is filled to a certain point, the tire walls resist further expansion, and the pressure increases as more air is added. (c) Once the tire is inflated fully, its pressure increases with temperature.

Pressure–Volume Work

Pressure–volume work is the work that is done by the compression or expansion of a fluid. Whenever there is a change in volume and external pressure remains constant, pressure–volume work is taking place. During a compression, a decrease in volume increases the internal pressure of a system as work is done *on* the system. During an expansion (Figure 12.5), an increase in volume decreases the internal pressure of a system as the system *does* work.

Figure 12.5 An expansion of a gas requires energy transfer to keep the pressure constant. Because pressure is constant, the work done is $P\Delta V$.

Recall that the formula for work is $W = Fd$. We can rearrange the definition of pressure, $P = \frac{F}{A}$, to get an expression for force in terms of pressure.

$$F = PA \tag{12.3}$$

Substituting this expression for force into the definition of work, we get

$$W = PAd. \tag{12.4}$$

Because area multiplied by displacement is the change in volume, $W = P\Delta V$, the mathematical expression for pressure–volume work is

$$W = P\Delta V. \tag{12.5}$$

Just as we say that work is force acting over a distance, for fluids, we can say that work is the pressure acting through the change in volume. For pressure–volume work, pressure is analogous to force, and volume is analogous to distance in the traditional definition of work.

WATCH PHYSICS

Work from Expansion

This video describes work from expansion (or pressure–volume work). Sal combines the equations $W = P\Delta V$ and $\Delta U = Q - W$ to get $\Delta U = Q - P\Delta V$.

Click to view content (https://www.openstax.org/l/28expansionWork)

GRASP CHECK

If the volume of a system increases while pressure remains constant, is the value of work done by the system $\boldsymbol{W}$ positive or negative? Will this increase or decrease the internal energy of the system?

a. Positive; internal energy will decrease
b. Positive; internal energy will increase
c. Negative; internal energy will decrease
d. Negative; internal energy will increase

The First Law of Thermodynamics

Heat (Q) and work (W) are the two ways to add or remove energy from a system. The processes are very different. Heat is driven

by temperature differences, while work involves a force exerted through a distance. Nevertheless, heat and work can produce identical results. For example, both can cause a temperature increase. Heat transfers energy into a system, such as when the sun warms the air in a bicycle tire and increases the air's temperature. Similarly, work can be done on the system, as when the bicyclist pumps air into the tire. Once the temperature increase has occurred, it is impossible to tell whether it was caused by heat or work. Heat and work are both energy in transit—neither is stored as such in a system. However, both can change the internal energy, U, of a system.

Internal energy is the sum of the kinetic and potential energies of a system's atoms and molecules. It can be divided into many subcategories, such as thermal and chemical energy, and depends only on the state of a system (that is, P, V, and T), not on how the energy enters or leaves the system.

In order to understand the relationship between heat, work, and internal energy, we use the **first law of thermodynamics**. The first law of thermodynamics applies the *conservation of energy* principle to systems where heat and work are the methods of transferring energy into and out of the systems. It can also be used to describe how energy transferred by heat is converted and transferred again by work.

TIPS FOR SUCCESS

Recall that the principle of conservation of energy states that energy cannot be created or destroyed, but it can be altered from one form to another.

The first law of thermodynamics states that the change in internal energy of a closed system equals the net heat transfer *into* the system minus the net work done *by* the system. In equation form, the first law of thermodynamics is

$$\Delta U = Q - W. \tag{12.6}$$

Here, ΔU is the *change in internal energy*, U, of the system. As shown in Figure 12.6, Q is the *net heat transferred into the system*—that is, Q is the sum of all heat transfers into and out of the system. W is the *net work done by the system*—that is, W is the sum of all work done on or by the system. By convention, if Q is positive, then there is a net heat transfer into the system; if W is positive, then there is net work done by the system. So positive Q adds energy to the system by heat, and positive W takes energy from the system by work. Note that if heat transfers more energy into the system than that which is done by work, the difference is stored as internal energy.

Figure 12.6 The first law of thermodynamics is the *conservation of energy* principle stated for a system, where heat and work are the methods of transferring energy to and from a system. Q represents the net heat transfer—it is the sum of all transfers of energy by heat into and out of the system. Q is positive for net heat transfer *into* the system. W_{out} is the work done *by* the system, and W_{in} is the work done *on* the system. W is the total work done on or *by* the system. W is positive when more work is done *by* the system than *on* it. The change in the internal energy of the system, ΔU, is related to heat and work by the first law of thermodynamics: $\Delta U = Q - W$.

It follows also that negative Q indicates that energy is transferred *away* from the system by heat and so decreases the system's internal energy, whereas negative W is work done *on* the system, which increases the internal energy.

WATCH PHYSICS

First Law of Thermodynamics/Internal Energy

This video explains the first law of thermodynamics, conservation of energy, and internal energy. It goes over an example of energy transforming between kinetic energy, potential energy, and heat transfer due to air resistance.

Click to view content (https://www.openstax.org/l/28FirstThermo)

GRASP CHECK

Consider the example of tossing a ball when there's air resistance. As air resistance increases, what would you expect to happen to the final velocity and final kinetic energy of the ball? Why?

a. Both will decrease. Energy is transferred to the air by heat due to air resistance.
b. Both will increase. Energy is transferred from the air to the ball due to air resistance.
c. Final velocity will increase, but final kinetic energy will decrease. Energy is transferred by heat to the air from the ball through air resistance.
d. Final velocity will decrease, but final kinetic energy will increase. Energy is transferred by heat from the air to the ball through air resistance.

WATCH PHYSICS

More on Internal Energy

This video goes into further detail, explaining internal energy and how to use the equation $\Delta U = Q - W$. Note that Sal uses the equation $\Delta U = Q + W$, where W is the work done *on* the system, whereas we use W to represent work done *by* the system.

Click to view content (https://www.openstax.org/l/28IntrnEnergy)

GRASP CHECK

If 5 J are taken away by heat from the system, and the system does 5 J of work, what is the change in internal energy of the system?

a. -10 J
b. 0 J
c. 10 J
d. 25 J

LINKS TO PHYSICS

Biology: Biological Thermodynamics

We often think about thermodynamics as being useful for inventing or testing machinery, such as engines or steam turbines. However, thermodynamics also applies to living systems, such as our own bodies. This forms the basis of the biological thermodynamics (Figure 12.7).

Figure 12.7 (a) The first law of thermodynamics applies to metabolism. Heat transferred out of the body (Q) and work done by the body (W) remove internal energy, whereas food intake replaces it. (Food intake may be considered work done on the body.) (b) Plants convert part of

the radiant energy in sunlight into stored chemical energy, a process called *photosynthesis*.

Life itself depends on the biological transfer of energy. Through photosynthesis, plants absorb solar energy from the sun and use this energy to convert carbon dioxide and water into glucose and oxygen. Photosynthesis takes in one form of energy—light—and converts it into another form—chemical potential energy (glucose and other carbohydrates).

Human metabolism is the conversion of food into energy given off by heat, work done by the body's cells, and stored fat. Metabolism is an interesting example of the first law of thermodynamics in action. Eating increases the internal energy of the body by adding chemical potential energy; this is an unromantic view of a good burrito.

The body metabolizes all the food we consume. Basically, metabolism is an oxidation process in which the chemical potential energy of food is released. This implies that food input is in the form of work. Exercise helps you lose weight, because it provides energy transfer from your body by both heat and work and raises your metabolic rate even when you are at rest.

Biological thermodynamics also involves the study of transductions between cells and living organisms. *Transduction* is a process where genetic material—DNA—is transferred from one cell to another. This often occurs during a viral infection (e.g., influenza) and is how the virus spreads, namely, by transferring its genetic material to an increasing number of previously healthy cells. Once enough cells become infected, you begin to feel the effects of the virus (flu symptoms—muscle weakness, coughing, and congestion).

Energy is transferred along with the genetic material and so obeys the first law of thermodynamics. Energy is transferred—not created or destroyed—in the process. When work is done on a cell or heat transfers energy to a cell, the cell's internal energy increases. When a cell does work or loses heat, its internal energy decreases. If the amount of work done by a cell is the same as the amount of energy transferred in by heat, or the amount of work performed on a cell matches the amount of energy transferred out by heat, there will be no net change in internal energy.

GRASP CHECK

Based on what you know about heat transfer and the first law of thermodynamics, do you need to eat more or less to maintain a constant weight in colder weather? Explain why.

a. more; as more energy is lost by the body in colder weather, the need to eat increases so as to maintain a constant weight
b. more; eating more food means accumulating more fat, which will insulate the body from colder weather and will reduce the energy loss
c. less; as less energy is lost by the body in colder weather, the need to eat decreases so as to maintain a constant weight
d. less; eating less food means accumulating less fat, so less energy will be required to burn the fat, and, as a result, weight will remain constant

Solving Problems Involving the First Law of Thermodynamics

WORKED EXAMPLE

Calculating Change in Internal Energy

Suppose 40.00 J of energy is transferred by heat to a system, while the system does 10.00 J of work. Later, heat transfers 25.00 J out of the system, while 4.00 J is done by work on the system. What is the net change in the system's internal energy?

STRATEGY

You must first calculate the net heat and net work. Then, using the first law of thermodynamics, $\Delta U = Q - W$, find the change in internal energy.

Solution

The net heat is the transfer into the system by heat minus the transfer out of the system by heat, or

$$Q = 40.00\text{ J} - 25.00\text{ J} = 15.00\text{ J}. \tag{12.7}$$

The total work is the work done by the system minus the work done on the system, or

$$W = 10.00\text{ J} - 4.00\text{ J} = 6.00\text{ J}. \tag{12.8}$$

The change in internal energy is given by the first law of thermodynamics.

$$\Delta U = Q - W = 15.00\text{ J} - 6.00\text{ J} = 9.00\text{ J} \tag{12.9}$$

Discussion

A different way to solve this problem is to find the change in internal energy for each of the two steps separately and then add the two changes to get the total change in internal energy. This approach would look as follows:

For 40.00 J of heat in and 10.00 J of work out, the change in internal energy is

$$\Delta U_1 = Q_1 - W_1 = 40.00\text{ J} - 10.00\text{ J} = 30.00\text{ J}. \tag{12.10}$$

For 25.00 J of heat out and 4.00 J of work in, the change in internal energy is

$$\Delta U_2 = Q_2 - W_2 = -25.\ 00\text{ J} - (-4.00\ \text{ J}) = -21.\ 00\text{ J}. \tag{12.11}$$

The total change is

$$\Delta U = \Delta U_1 + \Delta U_2 = 3\ 0.00\text{ J} + (-21.\ 00\text{ J}) = 9.\ 00\text{ J}. \tag{12.12}$$

No matter whether you look at the overall process or break it into steps, the change in internal energy is the same.

WORKED EXAMPLE

Calculating Change in Internal Energy: The Same Change in *U* is Produced by Two Different Processes

What is the change in the internal energy of a system when a total of 150.00 J is transferred by heat from the system and 159.00 J is done by work on the system?

STRATEGY

The net heat and work are already given, so simply use these values in the equation $\Delta U = Q - W$.

Solution

Here, the net heat and total work are given directly as $Q = -150.00\text{ J}$ and $W = -159.00\text{ J}$, so that

$$\Delta U = Q - W = -150.\ 00\text{ J} - (-159.\ 00\text{ J}) = 9.\ 00\text{ J}. \tag{12.13}$$

Discussion

Figure 12.8 Two different processes produce the same change in a system. (a) A total of 15.00 J of heat transfer occurs into the system, while work takes out a total of 6.00 J. The change in internal energy is ΔU = Q – W = 9.00 J. (b) Heat transfer removes 150.00 J from the system while work puts 159.00 J into it, producing an increase of 9.00 J in internal energy. If the system starts out in the same state in (a) and (b), it will end up in the same final state in either case—its final state is related to internal energy, not how that energy was acquired.

A very different process in this second worked example produces the same 9.00 J change in internal energy as in the first worked example. Note that the change in the system in both parts is related to ΔU and not to the individual Q's or W's involved. The system ends up in the *same* state in both problems. Note that, as usual, in Figure 12.8 above, W_{out} is work done *by* the system, and W_{in} is work done *on* the system.

Practice Problems

3. What is the pressure-volume work done by a system if a pressure of 20 Pa causes a change in volume of 3.0 m^3?
   a. 0.15 J
   b. 6.7 J
   c. 23 J
   d. 60 J

4. What is the net heat out of the system when 25 J is transferred by heat into the system and 45 J is transferred out of it?
   a. -70 J
   b. -20 J
   c. 20 J
   d. 70 J

Check Your Understanding

5. What is pressure?
   a. Pressure is force divided by length.
   b. Pressure is force divided by area.
   c. Pressure is force divided by volume.
   d. Pressure is force divided by mass.

6. What is the SI unit for pressure?
   a. pascal, or N/m^3
   b. coulomb
   c. newton
   d. pascal, or N/m^2

7. What is pressure-volume work?
   a. It is the work that is done by the compression or expansion of a fluid.
   b. It is the work that is done by a force on an object to produce a certain displacement.
   c. It is the work that is done by the surface molecules of a fluid.
   d. It is the work that is done by the high-energy molecules of a fluid.

8. When is pressure-volume work said to be done ON a system?
   a. When there is an increase in both volume and internal pressure.
   b. When there is a decrease in both volume and internal pressure.
   c. When there is a decrease in volume and an increase in internal pressure.
   d. When there is an increase in volume and a decrease in internal pressure.

9. What are the ways to add energy to or remove energy from a system?
   a. Transferring energy by heat is the only way to add energy to or remove energy from a system.
   b. Doing compression work is the only way to add energy to or remove energy from a system.
   c. Doing expansion work is the only way to add energy to or remove energy from a system.
   d. Transferring energy by heat or by doing work are the ways to add energy to or remove energy from a system.

10. What is internal energy?
   a. It is the sum of the kinetic energies of a system's atoms and molecules.
   b. It is the sum of the potential energies of a system's atoms and molecules.
   c. It is the sum of the kinetic and potential energies of a system's atoms and molecules.
   d. It is the difference between the magnitudes of the kinetic and potential energies of a system's atoms and molecules.

12.3 Second Law of Thermodynamics: Entropy

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Describe entropy
- Describe the second law of thermodynamics
- Solve problems involving the second law of thermodynamics

Section Key Terms

entropy	second law of thermodynamics

Entropy

Recall from the chapter introduction that it is not even theoretically possible for engines to be 100 percent efficient. This phenomenon is explained by the **second law of thermodynamics**, which relies on a concept known as **entropy**. Entropy is a measure of the disorder of a system. Entropy also describes how much energy is *not* available to do work. The more disordered a system and higher the entropy, the less of a system's energy is available to do work.

Although all forms of energy can be used to do work, it is not possible to use the entire available energy for work. Consequently, not all energy transferred by heat can be converted into work, and some of it is lost in the form of waste heat—that is, heat that does not go toward doing work. The unavailability of energy is important in thermodynamics; in fact, the field originated from efforts to convert heat to work, as is done by engines.

The equation for the change in entropy, ΔS, is

$$\Delta S = \frac{Q}{T},$$

where Q is the heat that transfers energy during a process, and T is the absolute temperature at which the process takes place.

Q is positive for energy transferred *into* the system by heat and negative for energy transferred *out of* the system by heat. In SI, entropy is expressed in units of joules per kelvin (J/K). If temperature changes during the process, then it is usually a good approximation (for small changes in temperature) to take T to be the average temperature in order to avoid trickier math (calculus).

TIPS FOR SUCCESS

Absolute temperature is the temperature measured in Kelvins. The Kelvin scale is an absolute temperature scale that is measured in terms of the number of degrees above absolute zero. All temperatures are therefore positive. Using temperatures from another, nonabsolute scale, such as Fahrenheit or Celsius, will give the wrong answer.

Second Law of Thermodynamics

Have you ever played the card game 52 pickup? If so, you have been on the receiving end of a practical joke and, in the process, learned a valuable lesson about the nature of the universe as described by the second law of thermodynamics. In the game of 52 pickup, the prankster tosses an entire deck of playing cards onto the floor, and you get to pick them up. In the process of picking up the cards, you may have noticed that the amount of work required to restore the cards to an orderly state in the deck is much greater than the amount of work required to toss the cards and create the disorder.

The second law of thermodynamics states that *the total entropy of a system either increases or remains constant in any spontaneous process; it never decreases.* An important implication of this law is that heat transfers energy spontaneously from higher- to lower-temperature objects, but never spontaneously in the reverse direction. This is because entropy increases for heat transfer of energy from hot to cold (Figure 12.9). Because the change in entropy is Q/T, there is a larger change in ΔS at lower temperatures (smaller T). The decrease in entropy of the hot (larger T) object is therefore less than the increase in entropy of the cold (smaller T) object, producing an overall increase in entropy for the system.

Figure 12.9 The ice in this drink is slowly melting. Eventually, the components of the liquid will reach thermal equilibrium, as predicted by the second law of thermodynamics—that is, after heat transfers energy from the warmer liquid to the colder ice. (Jon Sullivan, PDPhoto.org)

Another way of thinking about this is that it is impossible for any process to have, as its sole result, heat transferring energy from a cooler to a hotter object. Heat cannot transfer energy spontaneously from colder to hotter, because the entropy of the

overall system would decrease.

Suppose we mix equal masses of water that are originally at two different temperatures, say $20.0\ ^\circ\text{C}$ and $40.0\ ^\circ\text{C}$. The result will be water at an intermediate temperature of $30.0\ ^\circ\text{C}$. Three outcomes have resulted: entropy has increased, some energy has become unavailable to do work, and the system has become less orderly. Let us think about each of these results.

First, why has entropy increased? Mixing the two bodies of water has the same effect as the heat transfer of energy from the higher-temperature substance to the lower-temperature substance. The mixing decreases the entropy of the hotter water but increases the entropy of the colder water by a greater amount, producing an overall increase in entropy.

Second, once the two masses of water are mixed, there is no more temperature difference left to drive energy transfer by heat and therefore to do work. The energy is still in the water, but it is now *unavailable* to do work.

Third, the mixture is less orderly, or to use another term, less structured. Rather than having two masses at different temperatures and with different distributions of molecular speeds, we now have a single mass with a broad distribution of molecular speeds, the average of which yields an intermediate temperature.

These three results—entropy, unavailability of energy, and disorder—not only are related but are, in fact, essentially equivalent. Heat transfer of energy from hot to cold is related to the tendency in nature for systems to become disordered and for less energy to be available for use as work.

Based on this law, what cannot happen? A cold object in contact with a hot one never spontaneously transfers energy by heat to the hot object, getting colder while the hot object gets hotter. Nor does a hot, stationary automobile ever spontaneously cool off and start moving.

Another example is the expansion of a puff of gas introduced into one corner of a vacuum chamber. The gas expands to fill the chamber, but it never regroups on its own in the corner. The random motion of the gas molecules could take them all back to the corner, but this is never observed to happen (Figure 12.10).

Figure 12.10 Examples of one-way processes in nature. (a) Heat transfer occurs spontaneously from hot to cold, but not from cold to hot. (b) The brakes of this car convert its kinetic energy to increase their internal energy (temperature), and heat transfers this energy to the environment. The reverse process is impossible. (c) The burst of gas released into this vacuum chamber quickly expands to uniformly fill every part of the chamber. The random motions of the gas molecules will prevent them from returning altogether to the corner.

We've explained that heat never transfers energy spontaneously from a colder to a hotter object. The key word here is *spontaneously*. If we *do work* on a system, it *is* possible to transfer energy by heat from a colder to hotter object. We'll learn more about this in the next section, covering refrigerators as one of the applications of the laws of thermodynamics.

Sometimes people misunderstand the second law of thermodynamics, thinking that based on this law, it is impossible for entropy to decrease at any particular location. But, it actually *is* possible for the entropy of *one part* of the universe to decrease, as long as the total change in entropy of the universe increases. In equation form, we can write this as

$$\Delta S_{\text{tot}} = \Delta S_{\text{syst}} + \Delta S_{\text{envir}} > 0.$$

Based on this equation, we see that ΔS_{syst} can be negative as long as ΔS_{envir} is positive and greater in magnitude.

How is it possible for the entropy of a system to decrease? Energy transfer is necessary. If you pick up marbles that are scattered about the room and put them into a cup, your work has decreased the entropy of that system. If you gather iron ore from the ground and convert it into steel and build a bridge, your work has decreased the entropy of that system. Energy coming from the sun can decrease the entropy of local systems on Earth—that is, ΔS_{syst} is negative. But the overall entropy of the rest of the universe increases by a greater amount—that is, ΔS_{envir} is positive and greater in magnitude. In the case of the iron ore, although you made the system of the bridge and steel more structured, you did so at the expense of the universe. Altogether, the entropy of the universe is increased by the disorder created by digging up the ore and converting it to steel. Therefore,

$$\Delta S_{\text{tot}} = \Delta S_{\text{syst}} + \Delta S_{\text{envir}} > 0, \tag{12.14}$$

and the second law of thermodynamics is *not* violated.

Every time a plant stores some solar energy in the form of chemical potential energy, or an updraft of warm air lifts a soaring bird, Earth experiences local decreases in entropy as it uses part of the energy transfer from the sun into deep space to do work. There is a large total increase in entropy resulting from this massive energy transfer. A small part of this energy transfer by heat is stored in structured systems on Earth, resulting in much smaller, local decreases in entropy.

Solving Problems Involving the Second Law of Thermodynamics

Entropy is related not only to the unavailability of energy to do work; it is also a measure of disorder. For example, in the case of a melting block of ice, a highly structured and orderly system of water molecules changes into a disorderly liquid, in which molecules have no fixed positions (Figure 12.11). There is a large increase in entropy for this process, as we'll see in the following worked example.

Figure 12.11 These ice floes melt during the Arctic summer. Some of them refreeze in the winter, but the second law of thermodynamics predicts that it would be extremely unlikely for the water molecules contained in these particular floes to reform in the distinctive alligator-like shape they possessed when this picture was taken in the summer of 2009. (Patrick Kelley, U.S. Coast Guard, U.S. Geological Survey)

WORKED EXAMPLE

Entropy Associated with Disorder

Find the increase in entropy of 1.00 kg of ice that is originally at $0\ ^\circ\text{C}$ and melts to form water at $0\ ^\circ\text{C}$.

STRATEGY

The change in entropy can be calculated from the definition of ΔS once we find the energy, Q, needed to melt the ice.

Solution

The change in entropy is defined as

$$\Delta S = \frac{Q}{T}. \tag{12.15}$$

Here, Q is the heat necessary to melt 1.00 kg of ice and is given by

$$Q = mL_f, \tag{12.16}$$

where m is the mass and L_f is the latent heat of fusion. $L_f = 334\ \text{kJ/kg}$ for water, so

$$Q = (1.00\ \text{kg})(334\ \text{kJ/kg}) = 3.34 \times 10^5\,\text{J}. \tag{12.17}$$

Because Q is the amount of energy heat adds to the ice, its value is positive, and T is the melting temperature of ice, $T = 273\ \text{K}$ So the change in entropy is

$$\Delta S = \frac{Q}{T} = \frac{3.34 \times 10^5\,\text{J}}{273\ \text{K}} = 1.22 \times 10^3\,\text{J/K}. \tag{12.18}$$

Discussion

Figure 12.12 When ice melts, it becomes more disordered and less structured. The systematic arrangement of molecules in a crystal structure is replaced by a more random and less orderly movement of molecules without fixed locations or orientations. Its **entropy** increases because heat transfer occurs into it. Entropy is a measure of disorder.

The change in entropy is positive, because heat transfers energy *into* the ice to cause the phase change. This is a significant increase in entropy, because it takes place at a relatively low temperature. It is accompanied by an increase in the disorder of the water molecules.

Practice Problems

11. If $30.0\,\text{J}$ are added by heat to water at 12°C, what is the change in entropy?
- a. $0.105\,\text{J/K}$
- b. $2.5\,\text{J/K}$
- c. $0.45\,\text{J/K}$
- d. $9.50\,\text{J/K}$

12. What is the increase in entropy when $3.00\,\text{kg}$ of ice at 0°C melt to form water at 0°C?
- a. $1.84 \times 10^3\,\text{J/K}$
- b. $3.67 \times 10^3\,\text{J/K}$
- c. $1.84 \times 10^8\,\text{J/K}$
- d. $3.67 \times 10^8\,\text{J/K}$

Check Your Understanding

13. What is entropy?
- a. Entropy is a measure of the potential energy of a system.
- b. Entropy is a measure of the net work done by a system.
- c. Entropy is a measure of the disorder of a system.
- d. Entropy is a measure of the heat transfer of energy into a system.

14. Which forms of energy can be used to do work?
- a. Only work is able to do work.
- b. Only heat is able to do work.
- c. Only internal energy is able to do work.
- d. Heat, work, and internal energy are all able to do work.

15. What is the statement for the second law of thermodynamics?
- a. All the spontaneous processes result in decreased total entropy of a system.
- b. All the spontaneous processes result in increased total entropy of a system.
- c. All the spontaneous processes result in decreased or constant total entropy of a system.
- d. All the spontaneous processes result in increased or constant total entropy of a system.

16. For heat transferring energy from a high to a low temperature, what usually happens to the entropy of the whole system?
- a. It decreases.
- b. It must remain constant.
- c. The entropy of the system cannot be predicted without specific values for the temperatures.

d. It increases.

12.4 Applications of Thermodynamics: Heat Engines, Heat Pumps, and Refrigerators

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Explain how heat engines, heat pumps, and refrigerators work in terms of the laws of thermodynamics
- Describe thermal efficiency
- Solve problems involving thermal efficiency

Section Key Terms

cyclical process	heat engine	heat pump
thermal efficiency		

Heat Engines, Heat Pumps, and Refrigerators

In this section, we'll explore how heat engines, heat pumps, and refrigerators operate in terms of the laws of thermodynamics.

One of the most important things we can do with heat is to use it to do work for us. A **heat engine** does exactly this—it makes use of the properties of thermodynamics to transform heat into work. Gasoline and diesel engines, jet engines, and steam turbines that generate electricity are all examples of heat engines.

Figure 12.13 illustrates one of the ways in which heat transfers energy to do work. Fuel combustion releases chemical energy that heat transfers throughout the gas in a cylinder. This increases the gas temperature, which in turn increases the pressure of the gas and, therefore, the force it exerts on a movable piston. The gas does work on the outside world, as this force moves the piston through some distance. Thus, heat transfer of energy to the gas in the cylinder results in work being done.

Figure 12.13 (a) Heat transfer to the gas in a cylinder increases the internal energy of the gas, creating higher pressure and temperature. (b) The force exerted on the movable cylinder does work as the gas expands. Gas pressure and temperature decrease during expansion, indicating that the gas's internal energy has decreased as it does work. (c) Heat transfer of energy to the environment further reduces pressure in the gas, so that the piston can more easily return to its starting position.

To repeat this process, the piston needs to be returned to its starting point. Heat now transfers energy from the gas to the surroundings, so that the gas's pressure decreases, and a force is exerted by the surroundings to push the piston back through some distance.

A **cyclical process** brings a system, such as the gas in a cylinder, back to its original state at the end of every cycle. All heat engines use cyclical processes.

Heat engines do work by using part of the energy transferred by heat from some source. As shown in Figure 12.14, heat transfers energy, Q_h, from the high-temperature object (or hot reservoir), whereas heat transfers unused energy, Q_c, into the low-temperature object (or cold reservoir), and the work done by the engine is W. In physics, a *reservoir* is defined as an infinitely large mass that can take in or put out an unlimited amount of heat, depending upon the needs of the system. The temperature of the hot reservoir is T_h, and the temperature of the cold reservoir is T_c .

Figure 12.14 (a) Heat transfers energy spontaneously from a hot object to a cold one, as is consistent with the second law of thermodynamics. (b) A heat engine, represented here by a circle, uses part of the energy transferred by heat to do work. The hot and cold objects are called the hot and cold reservoirs. Q_h is the heat out of the hot reservoir, W is the work output, and Q_c is the unused heat into the cold reservoir.

As noted, a cyclical process brings the system back to its original condition at the end of every cycle. Such a system's internal energy, U, is the same at the beginning and end of every cycle—that is, $\Delta U = 0$. The first law of thermodynamics states that $\Delta U = Q - W$, where Q is the *net* heat transfer during the cycle, and W is the *net* work done by the system. The net heat transfer is the energy transferred in by heat from the hot reservoir minus the amount that is transferred out to the cold reservoir ($Q = Q_h - Q_c$). Because there is no change in internal energy for a complete cycle ($\Delta U = 0$), we have

$$0 = Q - W, \tag{12.19}$$

so that

$$W = Q. \tag{12.20}$$

Therefore, the net work done by the system equals the net heat into the system, or

$$W = Q_h - Q_c \tag{12.21}$$

for a cyclical process.

Because the hot reservoir is heated externally, which is an energy-intensive process, it is important that the work be done as efficiently as possible. In fact, we want W to equal Q_h, and for there to be no heat to the environment (that is, $Q_c = 0$). Unfortunately, this is impossible. According to the second law of thermodynamics, heat engines cannot have perfect conversion of heat into work. Recall that entropy is a measure of the disorder of a system, which is also how much energy is unavailable to do work. The second law of thermodynamics requires that the total entropy of a system either increases or remains constant in any process. Therefore, there is a minimum amount of Q_h that cannot be used for work. The amount of heat rejected to the cold reservoir, Q_c, depends upon the efficiency of the heat engine. The smaller the increase in entropy, ΔS, the smaller the value of Q_c, and the more heat energy is available to do work.

Heat pumps, air conditioners, and refrigerators utilize heat transfer of energy from low to high temperatures, which is the opposite of what heat engines do. Heat transfers energy Q_c from a cold reservoir and delivers energy Q_h into a hot one. This requires work input, W, which produces a transfer of energy by heat. Therefore, the total heat transfer to the hot reservoir is

$$Q_h = Q_c + W. \tag{12.22}$$

The purpose of a heat pump is to transfer energy by heat to a warm environment, such as a home in the winter. The great advantage of using a heat pump to keep your home warm rather than just burning fuel in a fireplace or furnace is that a heat pump supplies $Q_h = Q_c + W$. Heat Q_c comes from the outside air, even at a temperature below freezing, to the indoor space. You only pay for W, and you get an additional heat transfer of Q_c from the outside at no cost. In many cases, at least twice as much energy is transferred to the heated space as is used to run the heat pump. When you burn fuel to keep warm, you pay for all of it. The disadvantage to a heat pump is that the work input (required by the second law of thermodynamics) is sometimes

more expensive than simply burning fuel, especially if the work is provided by electrical energy.

The basic components of a heat pump are shown in Figure 12.15. A working fluid, such as a refrigerant, is used. In the outdoor coils (the evaporator), heat Q_c enters the working fluid from the cold outdoor air, turning it into a gas.

Figure 12.15 A simple heat pump has four basic components: (1) an evaporator, (2) a compressor, (3) a condenser, and (4) an expansion valve. In the heating mode, heat transfers Q_c to the working fluid in the evaporator (1) from the colder, outdoor air, turning it into a gas. The electrically driven compressor (2) increases the temperature and pressure of the gas and forces it into the condenser coils (3) inside the heated space. Because the temperature of the gas is higher than the temperature in the room, heat transfers energy from the gas to the room as the gas condenses into a liquid. The working fluid is then cooled as it flows back through an expansion valve (4) to the outdoor evaporator coils.

The electrically driven compressor (work input W) raises the temperature and pressure of the gas and forces it into the condenser coils that are inside the heated space. Because the temperature of the gas is higher than the temperature inside the room, heat transfers energy to the room, and the gas condenses into a liquid. The liquid then flows back through an expansion (pressure-reducing) valve. The liquid, having been cooled through expansion, returns to the outdoor evaporator coils to resume the cycle.

The quality of a heat pump is judged by how much energy is transferred by heat into the warm space (Q_h) compared with how much input work (W) is required.

Figure 12.16 Heat pumps, air conditioners, and refrigerators are heat engines operated backward. Almost every home contains a refrigerator. Most people don't realize that they are also sharing their homes with a heat pump.

Air conditioners and refrigerators are designed to cool substances by transferring energy by heat Q_c out of a cool environment to a warmer one, where heat Q_h is given up. In the case of a refrigerator, heat is moved out of the inside of the fridge into the

surrounding room. For an air conditioner, heat is transferred outdoors from inside a home. Heat pumps are also often used in a reverse setting to cool rooms in the summer.

As with heat pumps, work input is required for heat transfer of energy from cold to hot. The quality of air conditioners and refrigerators is judged by how much energy is removed by heat Q_c from a cold environment, compared with how much work, W, is required. So, what is considered the energy benefit in a heat pump, is considered waste heat in a refrigerator.

Thermal Efficiency

In the conversion of energy into work, we are always faced with the problem of getting less out than we put in. The problem is that, in all processes, there is some heat Q_c that transfers energy to the environment—and usually a very significant amount at that. A way to quantify how efficiently a machine runs is through a quantity called **thermal efficiency**.

We define thermal efficiency, *Eff*, to be the ratio of useful energy output to the energy input (or, in other words, the ratio of what we get to what we spend). The efficiency of a heat engine is the output of net work, W, divided by heat-transferred energy, Q_h, into the engine; that is

$$Eff = \frac{W}{Q_h}.$$

An efficiency of 1, or 100 percent, would be possible only if there were no heat to the environment ($Q_c = 0$).

TIPS FOR SUCCESS

All values of heat (Q_h and Q_c) are positive; there is no such thing as negative heat. The *direction* of heat is indicated by a plus or minus sign. For example, Q_c is out of the system, so it is preceded by a minus sign in the equation for net heat.

$$Q = Q_h - Q_c \tag{12.23}$$

Solving Thermal Efficiency Problems

WORKED EXAMPLE

Daily Work Done by a Coal-Fired Power Station and Its Efficiency

A coal-fired power station is a huge heat engine. It uses heat to transfer energy from burning coal to do work to turn turbines, which are used then to generate electricity. In a single day, a large coal power station transfers $2.50 \times 10^{14}\,\text{J}$ by heat from burning coal and transfers $1.48 \times 10^{14}\,\text{J}$ by heat into the environment. (a) What is the work done by the power station? (b) What is the efficiency of the power station?

STRATEGY

We can use $W = Q_h - Q_c$ to find the work output, W, assuming a cyclical process is used in the power station. In this process, water is boiled under pressure to form high-temperature steam, which is used to run steam turbine-generators and then condensed back to water to start the cycle again.

Solution

Work output is given by

$$W = Q_h - Q_c. \tag{12.24}$$

Substituting the given values,

$$W = 2.50 \times 10^{14}\,\text{J} - 1.48 \times 10^{14}\,\text{J} = 1.02 \times 10^{14}\,\text{J}. \tag{12.25}$$

STRATEGY

The efficiency can be calculated with $Eff = \frac{W}{Q_h}$, because Q_h is given, and work, W, was calculated in the first part of this example.

Solution

Efficiency is given by

$$Eff = \frac{W}{Q_{\text{h}}}. \tag{12.26}$$

The work, W, is found to be $1.02 \times 10^{14}\,\text{J}$, and Q_{h} is given ($2.50 \times 10^{14}\,\text{J}$), so the efficiency is

$$Eff = \frac{1.02 \times 10^{14}\,\text{J}}{2.50 \times 10^{14}\,\text{J}} = 0.408, \text{ or } 40.8\%. \tag{12.27}$$

Discussion

The efficiency found is close to the usual value of 42 percent for coal-burning power stations. It means that fully 59.2 percent of the energy is transferred by heat to the environment, which usually results in warming lakes, rivers, or the ocean near the power station and is implicated in a warming planet generally. While the laws of thermodynamics limit the efficiency of such plants—including plants fired by nuclear fuel, oil, and natural gas—the energy transferred by heat to the environment could be, and sometimes is, used for heating homes or for industrial processes.

Practice Problems

17. A heat engine is given $120\,\text{J}$ by heat and releases $20\,\text{J}$ by heat to the environment. What is the amount of work done by the system?

a. $-100\,\text{J}$
b. $-60\,\text{J}$
c. $60\,\text{J}$
d. $100\,\text{J}$

18. A heat engine takes in 6.0 kJ from heat and produces waste heat of 4.8 kJ. What is its efficiency?

a. 25 percent
b. 2.50 percent
c. 2.00 percent
d. 20 percent

Check Your Understanding

19. What is a heat engine?

a. A heat engine converts mechanical energy into thermal energy.
b. A heat engine converts thermal energy into mechanical energy.
c. A heat engine converts thermal energy into electrical energy.
d. A heat engine converts electrical energy into thermal energy.

20. Give an example of a heat engine.

a. A generator
b. A battery
c. A water pump
d. A car engine

21. What is thermal efficiency?

a. Thermal efficiency is the ratio of work input to the energy input.
b. Thermal efficiency is the ratio of work output to the energy input.
c. Thermal efficiency is the ratio of work input to the energy output.
d. Thermal efficiency is the ratio of work output to the energy output.

22. What is the mathematical expression for thermal efficiency?

a. $Eff = \frac{Q_{\text{h}}}{Q_{\text{h}} - Q_c}$
b. $Eff = \frac{Q_{\text{h}}}{Q_c}$
c. $Eff = \frac{Q_{\text{c}}}{Q_{\text{h}}}$
d. $Eff = \frac{Q_{\text{h}} - Q_{\text{c}}}{Q_{\text{h}}}$

KEY TERMS

Boltzmann constant constant with the value $k = 1.38 \times 10^{-23}$ J/K, which is used in the ideal gas law

cyclical process process in which a system is brought back to its original state at the end of every cycle

entropy measurement of a system's disorder and how much energy is not available to do work in a system

first law of thermodynamics states that the change in internal energy of a system equals the net energy transfer by heat *into* the system minus the net work done *by* the system

heat engine machine that uses energy transfer by heat to do work

heat pump machine that generates the heat transfer of energy from cold to hot

ideal gas law physical law that relates the pressure and volume of a gas to the number of gas molecules or atoms, or number of moles of gas, and the absolute temperature of the gas

internal energy sum of the kinetic and potential energies of a system's constituent particles (atoms or molecules)

pressure force per unit area perpendicular to the force, over which the force acts

second law of thermodynamics states that the total entropy of a system either increases or remains constant in any spontaneous process; it never decreases

thermal efficiency ratio of useful energy output to the energy input

thermal equilibrium condition in which heat no longer transfers energy between two objects that are in contact; the two objects have the same temperature

zeroth law of thermodynamics states that if two objects are in thermal equilibrium, and a third object is in thermal equilibrium with one of those objects, it is also in thermal equilibrium with the other object

SECTION SUMMARY

12.1 Zeroth Law of Thermodynamics: Thermal Equilibrium

- Systems are in thermal equilibrium when they have the same temperature.
- Thermal equilibrium occurs when two bodies are in contact with each other and can freely exchange energy.
- The zeroth law of thermodynamics states that when two systems, A and B, are in thermal equilibrium with each other, and B is in thermal equilibrium with a third system, C, then A is also in thermal equilibrium with C.

12.2 First law of Thermodynamics: Thermal Energy and Work

- Pressure is the force per unit area over which the force is applied perpendicular to the area.
- Thermal expansion is the increase, or decrease, of the size (length, area, or volume) of a body due to a change in temperature.
- The ideal gas law relates the pressure and volume of a gas to the number of gas particles (atoms or molecules) and the absolute temperature of the gas.
- Heat and work are the two distinct methods of energy transfer.
- Heat is energy transferred solely due to a temperature difference.
- The first law of thermodynamics is given as $\Delta U = Q - W$, where ΔU is the change in internal energy of a system, Q is the net energy transfer into the system by heat (the sum of all transfers by heat into and out of the system), and W is the net work done by the system (the sum of all energy transfers by work out of or into the system).
- Both Q and W represent energy in transit; only ΔU represents an independent quantity of energy capable of being stored.
- The internal energy U of a system depends only on the state of the system, and not how it reached that state.

12.3 Second Law of Thermodynamics: Entropy

- Entropy is a measure of a system's disorder: the greater the disorder, the larger the entropy.
- Entropy is also the reduced availability of energy to do work.
- The second law of thermodynamics states that, for any spontaneous process, the total entropy of a system either increases or remains constant; it never decreases.
- Heat transfers energy spontaneously from higher- to lower-temperature bodies, but never spontaneously in the reverse direction.

12.4 Applications of Thermodynamics: Heat Engines, Heat Pumps, and Refrigerators

- Heat engines use the heat transfer of energy to do work.
- Cyclical processes are processes that return to their original state at the end of every cycle.
- The thermal efficiency of a heat engine is the ratio of work output divided by the amount of energy input.
- The amount of work a heat engine can do is determined by the net heat transfer of energy during a cycle; more waste heat leads to less work output.

- Heat pumps draw energy by heat from cold outside air and use it to heat an interior room.
- A refrigerator is a type of heat pump; it takes energy from the warm air from the inside compartment and transfers it to warmer exterior air.

KEY EQUATIONS

12.2 First law of Thermodynamics: Thermal Energy and Work

ideal gas law	$PV = NkT$
first law of thermodynamics	$\Delta U = Q - W$
pressure	$P = \frac{F}{A}$
pressure–volume work	$W = P\Delta V$

12.3 Second Law of Thermodynamics: Entropy

change in entropy	$\Delta S = \frac{Q}{T}$

12.4 Applications of Thermodynamics: Heat Engines, Heat Pumps, and Refrigerators

thermal efficiency of a heat engine	$Eff = \frac{W}{Q_h}$
work output for a cyclical process	$W = Q_h - Q_c$

CHAPTER REVIEW

Concept Items

12.1 Zeroth Law of Thermodynamics: Thermal Equilibrium

1. When are two bodies in thermal equilibrium?
   a. When they are in thermal contact and are at different pressures
   b. When they are not in thermal contact but are at the same pressure
   c. When they are not in thermal contact but are at different temperatures
   d. When they are in thermal contact and are at the same temperature

2. What is thermal contact?
   a. Two objects are said to be in thermal contact when they are in contact with each other in such a way that the transfer of energy by heat can occur between them.
   b. Two objects are said to be in thermal contact when they are in contact with each other in such a way that the transfer of energy by mass can occur between them.
   c. Two objects are said to be in thermal contact when they neither lose nor gain energy by heat. There is no transfer of energy between them.
   d. Two objects are said to be in thermal contact when they only gain energy by heat. There is transfer of energy between them.

3. To which mathematical property is the zeroth law of thermodynamics similar?
   a. Associative property
   b. Commutative property
   c. Distributive property
   d. Transitive property

12.2 First law of Thermodynamics: Thermal Energy and Work

4. Why does thermal expansion occur?
   a. An increase in temperature causes intermolecular distances to increase.
   b. An increase in temperature causes intermolecular distances to decrease.
   c. An increase in temperature causes an increase in the work done on the system.
   d. An increase in temperature causes an increase in the work done by the system.

5. How does pressure-volume work relate to heat and internal energy of a system?
   a. The energy added to a system by heat minus the change in the internal energy of that system is equal to the pressure-volume work done by the system.
   b. The sum of the energy released by a system by heat and the change in the internal energy of that system is equal to the pressure-volume work done by the system.
   c. The product of the energy added to a system by heat and the change in the internal energy of that system

is equal to the pressure-volume work done by the system.

d. If the energy added to a system by heat is divided by the change in the internal energy of that system, the quotient is equal to the pressure-volume work done by the system.

6. On what does internal energy depend?
   a. The path of energy changes in the system
   b. The state of the system
   c. The size of the system
   d. The shape of the system

7. The first law of thermodynamics helps us understand the relationships among which three quantities?
   a. Heat, work, and internal energy
   b. Heat, work, and external energy
   c. Heat, work, and enthalpy
   d. Heat, work, and entropy

12.3 Second Law of Thermodynamics: Entropy

8. Air freshener is sprayed from a bottle. The molecules spread throughout the room and cannot make their way back into the bottle. Why is this the case?
   a. The entropy of the molecules increases.
   b. The entropy of the molecules decreases.
   c. The heat content (*enthalpy*, or total energy available for heat) of the molecules increases.
   d. The heat content (*enthalpy*, or total energy available for heat) of the molecules decreases.

9. Give an example of entropy as experienced in everyday life.
   a. rotation of Earth
   b. formation of a solar eclipse
   c. filling a car tire with air
   d. motion of a pendulum bob

12.4 Applications of Thermodynamics: Heat Engines, Heat Pumps, and Refrigerators

10. What is the quality by which air conditioners are judged?
   a. The amount of energy generated by heat from a hot environment, compared with the required work input
   b. The amount of energy transferred by heat from a cold environment, compared with the required work input
   c. The amount of energy transferred by heat from a hot environment, compared with the required work output
   d. The amount of energy transferred by heat from a cold environment, compared with the required work output

11. Why is the efficiency of a heat engine never 100 percent?
   a. Some energy is always gained by heat from the environment.
   b. Some energy is always lost by heat to the environment.
   c. Work output is always greater than energy input.
   d. Work output is infinite for any energy input.

12. What is a cyclic process?
   a. A process in which the system returns to its original state at the end of the cycle
   b. A process in which the system does not return to its original state at the end of the cycle
   c. A process in which the system follows the same path for every cycle
   d. A process in which the system follows a different path for every cycle

Critical Thinking Items

12.1 Zeroth Law of Thermodynamics: Thermal Equilibrium

13. What are the necessary conditions for energy transfer by heat to occur between two bodies through the process of conduction?
   a. They should be at the same temperature, and they should be in thermal contact.
   b. They should be at the same temperature, and they should not be in thermal contact.
   c. They should be at different temperatures, and they should be in thermal contact.
   d. They should be at different temperatures, and they should not be in thermal contact.

14. Oil is heated in a pan on a hot plate. The pan is in thermal equilibrium with the hot plate and also with the oil. The temperature of the hot plate is 150 °C . What is the temperature of the oil?
   a. 160 °C
   b. 150 °C
   c. 140 °C
   d. 130 °C

12.2 First law of Thermodynamics: Thermal Energy and Work

15. When an inflated balloon experiences a decrease in size,

the air pressure inside the balloon remains nearly constant. If there is no transfer of energy by heat to or from the balloon, what physical change takes place in the balloon?

a. The average kinetic energy of the gas particles decreases, so the balloon becomes colder.
b. The average kinetic energy of the gas particles increases, so the balloon becomes hotter.
c. The average potential energy of the gas particles decreases, so the balloon becomes colder.
d. The average potential energy of the gas particles increases, so the balloon becomes hotter.

16. When heat adds energy to a system, what is likely to happen to the pressure and volume of the system?

a. Pressure and volume may both decrease with added energy.
b. Pressure and volume may both increase with added energy.
c. Pressure must increase with added energy, while volume must remain constant.
d. Volume must decrease with added energy, while pressure must remain constant.

17. If more energy is transferred into the system by net heat as compared to the net work done by the system, what happens to the difference in energy?

a. It is transferred back to its surroundings.
b. It is stored in the system as internal energy.
c. It is stored in the system as potential energy.
d. It is stored in the system as entropy.

18. Air is pumped into a car tire, causing its temperature to increase. In another tire, the temperature increase is due to exposure to the sun. Is it possible to tell what caused the temperature increase in each tire? Explain your answer.

a. No, because it is a chemical change, and the cause of that change does not matter; the final state of both systems are the same.
b. Although the final state of each system is identical, the source is different in each case.
c. No, because the changes in energy for both systems are the same, and the cause of that change does not matter; the state of each system is identical.
d. Yes, the changes in the energy for both systems are the same, but the causes of that change are different, so the states of each system are not identical.

19. How does the transfer of energy from the sun help plants?

a. Plants absorb solar energy from the sun and utilize it during the fertilization process.
b. Plants absorb solar energy from the sun and utilize it during the process of photosynthesis to turn it into plant matter.
c. Plants absorb solar energy from the sun and utilize it to increase the temperature inside them.
d. Plants absorb solar energy from the sun and utilize it during the shedding of their leaves and fruits.

12.3 Second Law of Thermodynamics: Entropy

20. If an engine were constructed to perform such that there would be no losses due to friction, what would be its efficiency?

a. It would be 0 percent.
b. It would be less than 100 percent.
c. It would be 100 percent.
d. It would be greater than 100 percent.

21. Entropy never decreases in a spontaneous process. Give an example to support this statement.

a. The transfer of energy by heat from colder bodies to hotter bodies is a spontaneous process in which the entropy of the system of bodies increases.
b. The melting of an ice cube placed in a room causes an increase in the entropy of the room.
c. The dissolution of salt in water is a spontaneous process in which the entropy of the system increases.
d. A plant uses energy from the sun and converts it into sugar molecules by the process of photosynthesis.

12.4 Applications of Thermodynamics: Heat Engines, Heat Pumps, and Refrigerators

22. What is the advantage of a heat pump as opposed to burning fuel (as in a fireplace) for keeping warm?

a. A heat pump supplies energy by heat from the cold, outside air.
b. A heat pump supplies energy generated by the work done.
c. A heat pump supplies energy by heat from the cold, outside air and also from the energy generated by the work done.
d. A heat pump supplies energy not by heat from the cold, outside air, nor from the energy generated by the work done, but from more accessible sources.

23. What is thermal efficiency of an engine? Can it ever be 100 percent? Why or why not?

a. Thermal efficiency is the ratio of the output (work) to the input (heat). It is always 100 percent.
b. Thermal efficiency is the ratio of the output (heat)

to the input (work). It is always 100 percent.
c. Thermal efficiency is the ratio of the output (heat) to the input (work). It is never 100 percent.
d. Thermal efficiency is the ratio of the output (work) to the input (heat). It is never 100 percent.

24. When would 100 percent thermal efficiency be possible?
a. When all energy is transferred by heat to the environment
b. When mass transferred to the environment is zero
c. When mass transferred to the environment is at a maximum
d. When no energy is transferred by heat to the environment

Problems

12.2 First law of Thermodynamics: Thermal Energy and Work

25. Some amount of energy is transferred by heat into a system. The net work done by the system is $50\,\text{J}$, while the increase in its internal energy is $30\,\text{J}$. What is the amount of net heat?
a. $-80\,\text{J}$
b. $-20\,\text{J}$
c. $20\,\text{J}$
d. $80\,\text{J}$

26. Eighty joules are added by heat to a system, while it does $20\,\text{J}$ of work. Later, $30\,\text{J}$ are added by heat to the system, and it does $40\,\text{J}$ of work. What is the change in the system's internal energy?
a. $30\,\text{J}$
b. $50\,\text{J}$
c. $60\,\text{J}$
d. $110\,\text{J}$

12.4 Applications of Thermodynamics: Heat Engines, Heat Pumps, and Refrigerators

27. A coal power station functions at 40.0 percent efficiency. What is the amount of work it does if it takes in 1.20×10^{12} J by heat?
a. 3×10^{10} J
b. 4.8×10^{11} J
c. 3×10^{12} J
d. 4.8×10^{13} J

28. A heat engine functions with 70.7 percent thermal efficiency and consumes 12.0 kJ from heat daily. If its efficiency were raised to 75.0 percent, how much energy from heat would be saved daily, while providing the same output?
a. –10.8 kJ
b. –1.08 kJ
c. 0.7 kJ
d. 7 kJ

Performance Task

12.4 Applications of Thermodynamics: Heat Engines, Heat Pumps, and Refrigerators

29. You have been tasked to design and construct a thermometer that works on the principle of thermal expansion. There are four materials available for you to test, each of which will find use under different sets of conditions and temperature ranges:

Materials

- Four sample materials with similar mass or volume: copper, steel, water, and alcohol (ethanol or isopropanol)
- Oven or similar heating source
- Instrument (e.g., meter ruler, Vernier calipers, or micrometer) for measuring changes in dimension
- Balance for measuring mass

Procedure

1. Design a safe experiment to analyze the thermal expansion properties of each material.
2. Write down the materials needed for your experiment and the procedure you will follow. Make sure that you include every detail so that the experiment can be repeated by others.
3. Select an appropriate material to measure temperature over a predecided temperature range, and give reasons for your choice.
4. Calibrate your instrument to measure temperature changes accurately.

a. Which physical quantities are affected by temperature change and thermal expansion?
b. How do such properties as specific heat and thermal conductivity affect the use of each material as a thermometer?
c. Does a change of phase take place for any of the tested materials over the temperature range to be examined?
d. What are your independent and dependent variables for this series of tests? Which variables need to be controlled in the experiment?

e. What are your sources of error?
f. Can all the tested materials be used effectively in the same ranges of temperature? Which applications might be suitable for one or more of the tested substances but not the others?

TEST PREP

Multiple Choice

12.1 Zeroth Law of Thermodynamics: Thermal Equilibrium

30. Which law of thermodynamics describes thermal equilibrium?
a. zeroth
b. first
c. second
d. third

31. Name any two industries in which the principles of thermodynamics are used.
a. aerospace and information technology (IT) industries
b. industrial manufacturing and aerospace
c. mining and textile industries
d. mining and agriculture industries

12.2 First law of Thermodynamics: Thermal Energy and Work

32. What is the value of the Boltzmann constant?
a. $k = 1.23 \times 10^{-38}$ J/K
b. $k = 1.38 \times 10^{-23}$ J/K
c. $k = 1.38 \times 10^{23}$ J/K
d. $k = 1.23 \times 10^{38}$ J/K

33. Which of the following involves work done BY a system?
a. increasing internal energy
b. compression
c. expansion
d. cooling

34. Which principle does the first law of thermodynamics state?
a. the ideal gas law
b. the transitive property of equality
c. the law of conservation of energy
d. the principle of thermal equilibrium

35. What is the change in internal energy of a system when $Q_{\text{in}} = 50\,\text{J}$ and $Q_{\text{out}} = 50\,\text{J}$?
a. $20\,\text{J}$
b. $30\,\text{J}$
c. $50\,\text{J}$
d. $70\,\text{J}$

36. When does a real gas behave like an ideal gas?
a. A real gas behaves like an ideal gas at high temperature and low pressure.
b. A real gas behaves like an ideal gas at high temperature and high pressure.
c. A real gas behaves like an ideal gas at low temperature and low pressure.
d. A real gas behaves like an ideal gas at low temperature and high pressure.

12.3 Second Law of Thermodynamics: Entropy

37. In an engine, what is the unused energy converted into?
a. internal energy
b. pressure
c. work
d. heat

38. It is natural for systems in the universe to _____ spontaneously.
a. become disordered
b. become ordered
c. produce heat
d. do work

39. If Q is $120\,\text{J}$ and T is $350\,\text{K}$, what is the change in entropy?
a. $0.343\,\text{J/K}$
b. $1.51\,\text{J/K}$
c. $2.92\,\text{J/K}$
d. $34.3\,\text{J/K}$

40. Why does entropy increase during a spontaneous process?
a. Entropy increases because energy always transfers spontaneously from a dispersed state to a concentrated state.
b. Entropy increases because energy always transfers spontaneously from a concentrated state to a dispersed state.
c. Entropy increases because pressure always increases spontaneously.
d. Entropy increases because temperature of any system always increases spontaneously.

41. A system consists of ice melting in a glass of water. What happens to the entropy of this system?
a. The entropy of the ice decreases, while the entropy of the water cannot be predicted without more

specific information.
b. The entropy of the system remains constant.
c. The entropy of the system decreases.
d. The entropy of the system increases.

12.4 Applications of Thermodynamics: Heat Engines, Heat Pumps, and Refrigerators

42. Which equation represents the net work done by a system in a cyclic process?
a. $W = \frac{Q_c}{Q_h}$
b. $W = Q_h + Q_c$
c. $W = Eff\,(Q_c - Q_h)$
d. $W = Q_h - Q_c$

43. Which of these quantities needs to be zero for efficiency to be 100 percent?
a. ΔU
b. W
c. Q_h
d. Q_c

44. Which of the following always has the greatest value in a system having 80 percent thermal efficiency?
a. ΔU
b. W
c. Q_h
d. Q_c

45. In the equation $Q = Q_h - Q_c$, what does the negative sign indicate?
a. Heat transfer of energy is always negative.
b. Heat transfer can only occur in one direction.
c. Heat is directed into the system from the surroundings outside the system.
d. Heat is directed out of the system.

46. What is the purpose of a heat pump?
a. A heat pump uses work to transfer energy by heat from a colder environment to a warmer environment.
b. A heat pump uses work to transfer energy by heat from a warmer environment to a colder environment.
c. A heat pump does work by using heat to convey energy from a colder environment to a warmer environment.
d. A heat pump does work by using heat to convey energy from a warmer environment to a colder environment.

Short Answer

12.1 Zeroth Law of Thermodynamics: Thermal Equilibrium

47. What does *green* energy development entail?
a. Green energy involves finding new ways to harness clean and renewable alternative energy sources.
b. Green energy involves finding new ways to conserve alternative energy sources.
c. Green energy involves decreasing the efficiency of nonrenewable energy resources.
d. Green energy involves finding new ways to harness nonrenewable energy resources.

48. Why are the sun and Earth not in thermal equilibrium?
a. The mass of the sun is much greater than the mass of Earth.
b. There is a vast amount of empty space between the sun and Earth.
c. The diameter of the sun is much greater than the diameter of Earth.
d. The sun is in thermal contact with Earth.

12.2 First law of Thermodynamics: Thermal Energy and Work

49. If a fixed quantity of an ideal gas is held at a constant volume, which variable relates to pressure, and what is that relation?
a. Temperature; inverse proportionality $\left(P \propto \frac{1}{T}\right)$
b. Temperature, direct proportionality to square root $\left(P \propto \sqrt{T}\right)$
c. Temperature; direct proportionality $(P \propto T)$
d. Temperature; direct proportionality to square $\left(P \propto T^2\right)$

50. When is volume directly proportional to temperature?
a. when the pressure of the gas is variable
b. when the pressure of the gas is constant
c. when the mass of the gas is variable
d. when the mass of the gas is constant

51. For fluids, what can work be defined as?
a. pressure acting over the change in depth
b. pressure acting over the change in temperature
c. temperature acting over the change in volume
d. pressure acting over the change in volume

52. In the equation $\Delta U = Q - P\Delta V$, what does $P\Delta V$ indicate?
a. the work done on the system
b. the work done by the system
c. the heat into the system
d. the heat out of the system

53. By convention, if Q is positive, what is the direction in which heat transfers energy with regard to the system?
a. The direction of the heat transfer of energy depends on the changes in W, regardless of the sign of Q.
b. The direction of Q cannot be determined from just the sign of Q.
c. The direction of net heat transfer of energy will be out of the system.
d. The direction of net heat transfer of energy will be into the system.

54. What is net transfer of energy by heat?
a. It is the sum of all energy transfers by heat into the system.
b. It is the product of all energy transfers by heat into the system.
c. It is the sum of all energy transfers by heat into and out of the system.
d. It is the product of all energy transfers by heat into and out of the system.

55. Three hundred ten joules of heat enter a system, after which the system does 120 J of work. What is the change in its internal energy? Would this amount change if the energy transferred by heat were added after the work was done instead of before?
a. -190 J; this would change if heat added energy after the work was done
b. 190 J; this would change if heat added energy after the work was done
c. -190 J; this would not change even if heat added energy after the work was done
d. 190 J; this would not change even if heat added energy after the work was done

56. Ten joules are transferred by heat into a system, followed by another 20 J. What is the change in the system's internal energy? What would be the difference in this change if 30 J of energy were added by heat to the system at once?
a. 10 J; the change in internal energy would be same even if the heat added the energy at once
b. 30 J; the change in internal energy would be same even if the heat added the energy at once
c. 10 J; the change in internal energy would be more if the heat added the energy at once
d. 30 J; the change in internal energy would be more if the heat added the energy at once

12.3 Second Law of Thermodynamics: Entropy

57. How does the entropy of a system depend on how the system reaches a given state?
a. Entropy depends on the change of phase of a system, but not on any other state conditions.
b. Entropy does not depend on how the final state is reached from the initial state.
c. Entropy is least when the path between the initial state and the final state is the shortest.
d. Entropy is least when the path between the initial state and the final state is the longest.

58. Which sort of thermal energy do molecules in a solid possess?
a. electric potential energy
b. gravitational potential energy
c. translational kinetic energy
d. vibrational kinetic energy

59. A cold object in contact with a hot one never spontaneously transfers energy by heat to the hot object. Which law describes this phenomenon?
a. the first law of thermodynamics
b. the second law of thermodynamics
c. the third law of thermodynamics
d. the zeroth law of thermodynamics

60. How is it possible for us to transfer energy by heat from cold objects to hot ones?
a. by doing work on the system
b. by having work done by the system
c. by increasing the specific heat of the cold body
d. by increasing the specific heat of the hot body

61. What is the change in entropy caused by melting 5.00 kg of ice at 0 °C ?
a. 0 J/K
b. 6.11×10^3 J/K
c. 6.11×10^4 J/K
d. ∞J/K

62. What is the amount of heat required to cause a change of 35 J/K in the entropy of a system at 400 K?
a. $1.1 \times 10^1\text{ J}$
b. $1.1 \times 10^2\text{ J}$
c. $1.4 \times 10^3\text{ J}$
d. $1.4 \times 10^4\text{ J}$

12.4 Applications of Thermodynamics: Heat Engines, Heat Pumps, and Refrigerators

63. In a refrigerator, what is the function of an evaporator?
a. The evaporator converts gaseous refrigerant into liquid.
b. The evaporator converts solid refrigerant into liquid.
c. The evaporator converts solid refrigerant into gas.

d. The evaporator converts liquid refrigerant into gas.

64. Which component of an air conditioner converts gas into liquid?
a. the condenser
b. the compressor
c. the evaporator
d. the thermostat

65. What is one example for which calculating thermal efficiency is of interest?
a. A wind turbine
b. An electric pump
c. A bicycle
d. A car engine

66. How is the efficiency of a refrigerator or heat pump expressed?
a. $Eff = W\sqrt{Q_c}$
b. $Eff = \frac{W}{Q_c}$
c. $Eff = Q_c \times W$
d. $Eff = \frac{Q_c}{W}$

67. How can you mathematically express thermal efficiency in terms of Q_h and Q_c?
a. $Eff = \frac{Q_h}{Q_h - Q_c}$
b. $Eff = \frac{Q_h}{Q_c}$
c. $Eff = \frac{Q_c}{Q_h}$
d. $Eff = \frac{Q_h - Q_c}{Q_h}$

68. How can you calculate percentage efficiency?
a. percentage efficiency $= (Eff + 100)\ \%$
b. percentage efficiency $= \frac{Eff}{100}\%$
c. percentage efficiency $= (Eff - 100)\ \%$
d. percentage efficiency $= Eff \times 100\%$

Extended Response

12.1 Zeroth Law of Thermodynamics: Thermal Equilibrium

69. What is the meaning of efficiency in terms of a car engine?
a. An engine's efficiency equals the sum of useful energy (work) and the input energy.
b. An engine's efficiency equals the proportion of useful energy (work) to the input energy.
c. An engine's efficiency equals the product of useful energy (work) and the input energy.
d. An engine's efficiency equals the difference between the useful energy (work) and the input energy.

12.2 First law of Thermodynamics: Thermal Energy and Work

70. Why does a bridge have expansion joints?
a. because the bridge expands and contracts with the change in temperature
b. because the bridge expands and contracts with the change in motion of objects moving on the bridge
c. because the bridge expands and contracts with the change in total load on the bridge
d. because the bridge expands and contracts with the change in magnitude of wind blowing

71. Under which conditions will the work done by the gas in a system increase?
a. It will increase when a large amount of energy is added to the system, and that energy causes an increase in the gas's volume, its pressure, or both.
b. It will increase when a large amount of energy is extracted from the system, and that energy causes an increase in the gas's volume, its pressure, or both.
c. It will increase when a large amount of energy is added to the system, and that energy causes a decrease in the gas's volume, its pressure, or both.
d. It will increase when a large amount of energy is extracted from the system, and that energy causes a decrease in the gas's volume, its pressure, or both.

72. How does energy transfer by heat aid in body metabolism?
a. The energy is given to the body through the work done by the body (W) and through the intake of food, which may also be considered as the work done on the body. The transfer of energy out of the body is by heat ($-Q$) .
b. The energy given to the body is by the intake of food, which may also be considered as the work done on the body. The transfer of energy out of the body is by heat ($-Q$) and the work done by the body (W) .
c. The energy given to the body is by the transfer of energy by heat (Q) into the body, which may also be considered as the work done on the body. The transfer of energy out of the body is the work done by the body (W) .
d. The energy given to the body is by the transfer of energy by heat (Q) inside the body. The transfer of energy out of the body is by the intake of food and the work done by the body (W) .

73. Two distinct systems have the same amount of stored internal energy. Five hundred joules are added by heat to

the first system, and 300 J are added by heat to the second system. What will be the change in internal energy of the first system if it does 200 J of work? How much work will the second system have to do in order to have the same internal energy?

a. 700 J; 0 J
b. 300 J; 300 J
c. 700 J; 300 J
d. 300 J; 0 J

12.3 Second Law of Thermodynamics: Entropy

74. Why is it not possible to convert all available energy into work?

a. Due to the entropy of a system, some energy is always unavailable for work.
b. Due to the entropy of a system, some energy is always available for work.
c. Due to the decrease in internal energy of a system, some energy is always made unavailable for work.
d. Due to the increase in internal energy of a system, some energy is always made unavailable for work.

75. Why does entropy increase when ice melts into water?

a. Melting converts the highly ordered solid structure into a disorderly liquid, thereby increasing entropy.
b. Melting converts the highly ordered liquid into a disorderly solid structure, thereby increasing entropy.
c. Melting converts the highly ordered solid structure into a disorderly solid structure, thereby increasing entropy.
d. Melting converts the highly ordered liquid into a disorderly liquid, thereby increasing entropy.

76. Why is change in entropy lower for higher temperatures?

a. Increase in the disorder in the substance is low for high temperature.
b. Increase in the disorder in the substance is high for high temperature.
c. Decrease in the disorder in the substance is low for high temperature.
d. Decrease in the disorder in the substance is high for high temperature.

12.4 Applications of Thermodynamics: Heat Engines, Heat Pumps, and Refrigerators

77. In the equation $W = Q_h - Q_c$, if the value of Q_c were equal to zero, what would it signify?

a. The efficiency of the engine is 75 percent.
b. The efficiency of the engine is 25 percent.
c. The efficiency of the engine is 0 percent.
d. The efficiency of the engine is 100 percent.

78. Can the value of thermal efficiency be greater than 1? Why or why not?

a. No, according to the first law of thermodynamics, energy output can never be more than the energy input.
b. No, according to the second law of thermodynamics, energy output can never be more than the energy input.
c. Yes, according to the first law of thermodynamics, energy output can be more than the energy input.
d. Yes, according to the second law of thermodynamics, energy output can be more than the energy input.

79. A coal power station transfers 3.0×10^{12} J by heat from burning coal and transfers 1.5×10^{12} J by heat into the environment. What is the efficiency of the power station?

a. 0.33
b. 0.5
c. 0.66
d. 1

CHAPTER 13
Waves and Their Properties

Figure 13.1 Waves in the ocean behave similarly to all other types of waves. (Steve Jurveston, Flickr)

Chapter Outline

13.1 Types of Waves

13.2 Wave Properties: Speed, Amplitude, Frequency, and Period

13.3 Wave Interaction: Superposition and Interference

INTRODUCTION Recall from the chapter on Motion in Two Dimensions that oscillations—the back-and-forth movement between two points—involve force and energy. Some oscillations create waves, such as the sound waves created by plucking a guitar string. Other examples of waves include earthquakes and visible light. Even subatomic particles, such as electrons, can behave like waves. You can make water waves in a swimming pool by slapping the water with your hand. Some of these waves, such as water waves, are visible; others, such as sound waves, are not. But every wave is a disturbance that moves from its source and carries energy. In this chapter, we will learn about the different types of waves, their properties, and how they interact with one another.

13.1 Types of Waves

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Define mechanical waves and medium, and relate the two
- Distinguish a pulse wave from a periodic wave
- Distinguish a longitudinal wave from a transverse wave and give examples of such waves

Section Key Terms

longitudinal wave	mechanical wave	medium	wave
periodic wave	pulse wave	transverse wave	

Mechanical Waves

What do we mean when we say something is a wave? A **wave** is a disturbance that travels or *propagates* from the place where it was created. Waves transfer energy from one place to another, but they do not necessarily transfer any mass. Light, sound, and waves in the ocean are common examples of waves. Sound and water waves are **mechanical waves**; meaning, they require a medium to travel through. The **medium** may be a solid, a liquid, or a gas, and the speed of the wave depends on the material properties of the medium through which it is traveling. However, light is not a mechanical wave; it can travel through a vacuum such as the empty parts of outer space.

A familiar wave that you can easily imagine is the water wave. For water waves, the disturbance is in the surface of the water, an example of which is the disturbance created by a rock thrown into a pond or by a swimmer splashing the water surface repeatedly. For sound waves, the disturbance is caused by a change in air pressure, an example of which is when the oscillating cone inside a speaker creates a disturbance. For earthquakes, there are several types of disturbances, which include the disturbance of Earth's surface itself and the pressure disturbances under the surface. Even radio waves are most easily understood using an analogy with water waves. Because water waves are common and visible, visualizing water waves may help you in studying other types of waves, especially those that are not visible.

Water waves have characteristics common to all waves, such as amplitude, period, frequency, and energy, which we will discuss in the next section.

Pulse Waves and Periodic Waves

If you drop a pebble into the water, only a few waves may be generated before the disturbance dies down, whereas in a wave pool, the waves are continuous. A **pulse wave** is a sudden disturbance in which only one wave or a few waves are generated, such as in the example of the pebble. Thunder and explosions also create pulse waves. A **periodic wave** repeats the same oscillation for several cycles, such as in the case of the wave pool, and is associated with simple harmonic motion. Each particle in the medium experiences simple harmonic motion in periodic waves by moving back and forth periodically through the same positions.

Consider the simplified water wave in Figure 13.2. This wave is an up-and-down disturbance of the water surface, characterized by a sine wave pattern. The uppermost position is called the *crest* and the lowest is the *trough*. It causes a seagull to move up and down in simple harmonic motion as the wave crests and troughs pass under the bird.

Figure 13.2 An idealized ocean wave passes under a seagull that bobs up and down in simple harmonic motion.

Longitudinal Waves and Transverse Waves

Mechanical waves are categorized by their type of motion and fall into any of two categories: transverse or longitudinal. Note that both transverse and longitudinal waves can be periodic. A **transverse wave** propagates so that the disturbance is perpendicular to the direction of propagation. An example of a transverse wave is shown in Figure 13.3, where a woman moves a toy spring up and down, generating waves that propagate away from herself in the horizontal direction while disturbing the toy spring in the vertical direction.

Figure 13.3 In this example of a transverse wave, the wave propagates horizontally and the disturbance in the toy spring is in the vertical direction.

In contrast, in a **longitudinal wave**, the disturbance is parallel to the direction of propagation. Figure 13.4 shows an example of a longitudinal wave, where the woman now creates a disturbance in the horizontal direction—which is the same direction as the wave propagation—by stretching and then compressing the toy spring.

Figure 13.4 In this example of a longitudinal wave, the wave propagates horizontally and the disturbance in the toy spring is also in the horizontal direction.

TIPS FOR SUCCESS

Longitudinal waves are sometimes called *compression waves* or *compressional waves*, and transverse waves are sometimes called *shear waves*.

Waves may be transverse, longitudinal, or *a combination of the two*. The waves on the strings of musical instruments are transverse (as shown in Figure 13.5), and so are electromagnetic waves, such as visible light. Sound waves in air and water are longitudinal. Their disturbances are periodic variations in pressure that are transmitted in fluids.

Figure 13.5 The wave on a guitar string is transverse. However, the sound wave coming out of a speaker rattles a sheet of paper in a direction that shows that such sound wave is longitudinal.

Sound in solids can be both longitudinal and transverse. Essentially, water waves are also a combination of transverse and longitudinal components, although the simplified water wave illustrated in Figure 13.2 does not show the longitudinal motion of the bird.

Earthquake waves under Earth's surface have both longitudinal and transverse components as well. The longitudinal waves in an earthquake are called pressure or P-waves, and the transverse waves are called shear or S-waves. These components have important individual characteristics; for example, they propagate at different speeds. Earthquakes also have surface waves that are similar to surface waves on water.

WATCH PHYSICS

Introduction to Waves

This video explains wave propagation in terms of momentum using an example of a wave moving along a rope. It also covers the differences between transverse and longitudinal waves, and between pulse and periodic waves.

Click to view content (https://openstax.org/l/02introtowaves)

GRASP CHECK

In a longitudinal sound wave, after a compression wave moves through a region, the density of molecules briefly decreases. Why is this?

a. After a compression wave, some molecules move forward temporarily.
b. After a compression wave, some molecules move backward temporarily.
c. After a compression wave, some molecules move upward temporarily.
d. After a compression wave, some molecules move downward temporarily.

FUN IN PHYSICS

The Physics of Surfing

Many people enjoy surfing in the ocean. For some surfers, the bigger the wave, the better. In one area off the coast of central California, waves can reach heights of up to 50 feet in certain times of the year (Figure 13.6).

Figure 13.6 A surfer negotiates a steep take-off on a winter day in California while his friend watches. (Ljsurf, Wikimedia Commons)

How do waves reach such extreme heights? Other than unusual causes, such as when earthquakes produce tsunami waves, most huge waves are caused simply by interactions between the wind and the surface of the water. The wind pushes up against the surface of the water and transfers energy to the water in the process. The stronger the wind, the more energy transferred. As waves start to form, a larger surface area becomes in contact with the wind, and even more energy is transferred from the wind to the water, thus creating higher waves. Intense storms create the fastest winds, kicking up massive waves that travel out from the origin of the storm. Longer-lasting storms and those storms that affect a larger area of the ocean create the biggest waves since they transfer more energy. The cycle of the tides from the Moon's gravitational pull also plays a small role in creating waves.

Actual ocean waves are more complicated than the idealized model of the simple transverse wave with a perfect sinusoidal shape. Ocean waves are examples of *orbital progressive waves*, where water particles at the surface follow a circular path from the crest to the trough of the passing wave, then cycle back again to their original position. This cycle repeats with each passing wave.

As waves reach shore, the water depth decreases and the energy of the wave is compressed into a smaller volume. This creates higher waves—an effect known as *shoaling*.

Since the water particles along the surface move from the crest to the trough, surfers hitch a ride on the cascading water, gliding along the surface. If ocean waves work exactly like the idealized transverse waves, surfing would be much less exciting as it would simply involve standing on a board that bobs up and down in place, just like the seagull in the previous figure.

Additional information and illustrations about the scientific principles behind surfing can be found in the "Using Science to Surf Better!" (http://www.openstax.org/l/28Surf) video.

GRASP CHECK

If we lived in a parallel universe where ocean waves were longitudinal, what would a surfer's motion look like?

a. The surfer would move side-to-side/back-and-forth vertically with no horizontal motion.
b. The surfer would forward and backward horizontally with no vertical motion.

Check Your Understanding

1. What is a wave?
   a. A wave is a force that propagates from the place where it was created.
   b. A wave is a disturbance that propagates from the place where it was created.
   c. A wave is matter that provides volume to an object.
   d. A wave is matter that provides mass to an object.

2. Do all waves require a medium to travel? Explain.
   a. No, electromagnetic waves do not require any medium to propagate.
   b. No, mechanical waves do not require any medium to propagate.
   c. Yes, both mechanical and electromagnetic waves require a medium to propagate.
   d. Yes, all transverse waves require a medium to travel.

3. What is a pulse wave?
   a. A pulse wave is a sudden disturbance with only one wave generated.
   b. A pulse wave is a sudden disturbance with only one or a few waves generated.

c. A pulse wave is a gradual disturbance with only one or a few waves generated.
d. A pulse wave is a gradual disturbance with only one wave generated.

4. Is the following statement true or false? A pebble dropped in water is an example of a pulse wave.
   a. False
   b. True

5. What are the categories of mechanical waves based on the type of motion?
   a. Both transverse and longitudinal waves
   b. Only longitudinal waves
   c. Only transverse waves
   d. Only surface waves

6. In which direction do the particles of the medium oscillate in a transverse wave?
   a. Perpendicular to the direction of propagation of the transverse wave
   b. Parallel to the direction of propagation of the transverse wave

13.2 Wave Properties: Speed, Amplitude, Frequency, and Period

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Define amplitude, frequency, period, wavelength, and velocity of a wave
- Relate wave frequency, period, wavelength, and velocity
- Solve problems involving wave properties

Section Key Terms

wavelength	wave velocity

Wave Variables

In the chapter on motion in two dimensions, we defined the following variables to describe harmonic motion:

- Amplitude—maximum displacement from the equilibrium position of an object oscillating around such equilibrium position
- Frequency—number of events per unit of time
- Period—time it takes to complete one oscillation

For waves, these variables have the same basic meaning. However, it is helpful to word the definitions in a more specific way that applies directly to waves:

- Amplitude—distance between the resting position and the maximum displacement of the wave
- Frequency—number of waves passing by a specific point per second
- Period—time it takes for one wave cycle to complete

In addition to amplitude, frequency, and period, their wavelength and wave velocity also characterize waves. The **wavelength** λ is the distance between adjacent identical parts of a wave, parallel to the direction of propagation. The **wave velocity** v_w is the speed at which the disturbance moves.

TIPS FOR SUCCESS

Wave velocity is sometimes also called the *propagation velocity* or *propagation speed* because the disturbance propagates from one location to another.

Consider the periodic water wave in Figure 13.7. Its wavelength is the distance from crest to crest or from trough to trough. The wavelength can also be thought of as the distance a wave has traveled after one complete cycle—or one period. The time for one complete up-and-down motion is the simple water wave's period T. In the figure, the wave itself moves to the right with a wave

velocity v_w. Its amplitude X is the distance between the resting position and the maximum displacement—either the crest or the trough—of the wave. It is important to note that this movement of the wave is actually the *disturbance* moving to the right, not the water itself; otherwise, the bird would move to the right. Instead, the seagull bobs up and down in place as waves pass underneath, traveling a total distance of $2X$ in one cycle. However, as mentioned in the text feature on surfing, actual ocean waves are more complex than this simplified example.

Figure 13.7 The wave has a wavelength λ, which is the distance between adjacent identical parts of the wave. The up-and-down disturbance of the surface propagates parallel to the surface at a speed v_w.

WATCH PHYSICS

Amplitude, Period, Frequency, and Wavelength of Periodic Waves

This video is a continuation of the video "Introduction to Waves" from the "Types of Waves" section. It discusses the properties of a periodic wave: amplitude, period, frequency, wavelength, and wave velocity.

Click to view content (https://www.openstax.org/l/28wavepro)

TIPS FOR SUCCESS

The crest of a wave is sometimes also called the *peak*.

GRASP CHECK

If you are on a boat in the trough of a wave on the ocean, and the wave amplitude is 1 m, what is the wave height from your position?

a. 1 m
b. 2 m
c. 4 m
d. 8 m

The Relationship between Wave Frequency, Period, Wavelength, and Velocity

Since wave frequency is the number of waves per second, and the period is essentially the number of seconds per wave, the relationship between frequency and period is

$$f = \frac{1}{T} \tag{13.1}$$

or

$$T = \frac{1}{f}, \tag{13.2}$$

just as in the case of harmonic motion of an object. We can see from this relationship that a higher frequency means a shorter period. Recall that the unit for frequency is hertz (Hz), and that 1 Hz is one cycle—or one wave—per second.

The speed of propagation v_w is the distance the wave travels in a given time, which is one wavelength in a time of one period. In

equation form, it is written as

$$v_w = \frac{\lambda}{T} \tag{13.3}$$

or

$$v_w = f\lambda. \tag{13.4}$$

From this relationship, we see that in a medium where v_w is constant, the higher the frequency, the smaller the wavelength. See Figure 13.8.

Figure 13.8 Because they travel at the same speed in a given medium, low-frequency sounds must have a greater wavelength than high-frequency sounds. Here, the lower-frequency sounds are emitted by the large speaker, called a woofer, while the higher-frequency sounds are emitted by the small speaker, called a tweeter.

These fundamental relationships hold true for all types of waves. As an example, for water waves, v_w is the speed of a surface wave; for sound, v_w is the speed of sound; and for visible light, v_w is the speed of light. The amplitude X is completely independent of the speed of propagation v_w and depends only on the amount of energy in the wave.

Snap Lab

Waves in a Bowl

In this lab, you will take measurements to determine how the amplitude and the period of waves are affected by the transfer of energy from a cork dropped into the water. The cork initially has some potential energy when it is held above the water—the greater the height, the higher the potential energy. When it is dropped, such potential energy is converted to kinetic energy as the cork falls. When the cork hits the water, that energy travels through the water in waves.

- Large bowl or basin
- Water
- Cork (or ping pong ball)
- Stopwatch
- Measuring tape

Instructions

Procedure

1. Fill a large bowl or basin with water and wait for the water to settle so there are no ripples.
2. Gently drop a cork into the middle of the bowl.
3. Estimate the wavelength and the period of oscillation of the water wave that propagates away from the cork. You can estimate the period by counting the number of ripples from the center to the edge of the bowl while your partner times it. This information, combined with the bowl measurement, will give you the wavelength when the correct formula is used.
4. Remove the cork from the bowl and wait for the water to settle again.
5. Gently drop the cork at a height that is different from the first drop.
6. Repeat Steps 3 to 5 to collect a second and third set of data, dropping the cork from different heights and recording the resulting wavelengths and periods.

7. Interpret your results.

GRASP CHECK

A cork is dropped into a pool of water creating waves. Does the wavelength depend upon the height above the water from which the cork is dropped?

a. No, only the amplitude is affected.
b. Yes, the wavelength is affected.

LINKS TO PHYSICS

Geology: Physics of Seismic Waves

Figure 13.9 The destructive effect of an earthquake is a palpable evidence of the energy carried in the earthquake waves. The Richter scale rating of earthquakes is related to both their amplitude and the energy they carry. (Petty Officer 2nd Class Candice Villarreal, U.S. Navy)

Geologists rely heavily on physics to study earthquakes since earthquakes involve several types of wave disturbances, including disturbance of Earth's surface and pressure disturbances under the surface. Surface earthquake waves are similar to surface waves on water. The waves under Earth's surface have both longitudinal and transverse components. The longitudinal waves in an earthquake are called pressure waves (P-waves) and the transverse waves are called shear waves (S-waves). These two types of waves propagate at different speeds, and the speed at which they travel depends on the rigidity of the medium through which they are traveling. During earthquakes, the speed of P-waves in granite is significantly higher than the speed of S-waves. Both components of earthquakes travel more slowly in less rigid materials, such as sediments. P-waves have speeds of 4 to 7 km/s, and S-waves have speeds of 2 to 5 km/s, but both are faster in more rigid materials. The P-wave gets progressively farther ahead of the S-wave as they travel through Earth's crust. For that reason, the time difference between the P- and S-waves is used to determine the distance to their source, the epicenter of the earthquake.

We know from seismic waves produced by earthquakes that parts of the interior of Earth are liquid. Shear or transverse waves cannot travel through a liquid and are not transmitted through Earth's core. In contrast, compression or longitudinal waves can pass through a liquid and they do go through the core.

All waves carry energy, and the energy of earthquake waves is easy to observe based on the amount of damage left behind after the ground has stopped moving. Earthquakes can shake whole cities to the ground, performing the work of thousands of wrecking balls. The amount of energy in a wave is related to its amplitude. Large-amplitude earthquakes produce large ground displacements and greater damage. As earthquake waves spread out, their amplitude decreases, so there is less damage the farther they get from the source.

GRASP CHECK

What is the relationship between the propagation speed, frequency, and wavelength of the S-waves in an earthquake?

a. The relationship between the propagation speed, frequency, and wavelength is $v_w = f\lambda$.
b. The relationship between the propagation speed, frequency, and wavelength is $v_w = \frac{f}{\lambda}$.
c. The relationship between the propagation speed, frequency, and wavelength is $v_w = \frac{\lambda}{f}$.

d. The relationship between the propagation speed, frequency, and wavelength is $v_w = \sqrt{f\lambda}$.

Virtual Physics

Wave on a String

Click to view content (http://www.openstax.org/l/28wavestring)

In this animation, watch how a string vibrates in slow motion by choosing the Slow Motion setting. Select the No End and Manual options, and wiggle the end of the string to make waves yourself. Then switch to the Oscillate setting to generate waves automatically. Adjust the frequency and the amplitude of the oscillations to see what happens. Then experiment with adjusting the damping and the tension.

GRASP CHECK

Which of the settings—amplitude, frequency, damping, or tension—changes the amplitude of the wave as it propagates? What does it do to the amplitude?

a. Frequency; it decreases the amplitude of the wave as it propagates.
b. Frequency; it increases the amplitude of the wave as it propagates.
c. Damping; it decreases the amplitude of the wave as it propagates.
d. Damping; it increases the amplitude of the wave as it propagates.

Solving Wave Problems

WORKED EXAMPLE

Calculate the Velocity of Wave Propagation: Gull in the Ocean

Calculate the wave velocity of the ocean wave in the previous figure if the distance between wave crests is 10.0 m and the time for a seagull to bob up and down is 5.00 s.

STRATEGY

The values for the wavelength ($\lambda = 10.0\ \ m$) and the period ($T = 5.00$ s) are given and we are asked to find v_w Therefore, we can use $v_w = \frac{\lambda}{T}$ to find the wave velocity.

Solution

Enter the known values into $v_w = \frac{\lambda}{T}$

$$v_w = \frac{10.0\text{ m}}{5.00\text{ s}} = 2.\ 00\text{ m/s}. \tag{13.5}$$

Discussion

This slow speed seems reasonable for an ocean wave. Note that in the figure, the wave moves to the right at this speed, which is different from the varying speed at which the seagull bobs up and down.

WORKED EXAMPLE

Calculate the Period and the Wave Velocity of a Toy Spring

The woman in creates two waves every second by shaking the toy spring up and down. (a)What is the period of each wave? (b) If each wave travels 0.9 meters after one complete wave cycle, what is the velocity of wave propagation?

STRATEGY FOR (A)

To find the period, we solve for $T = \frac{1}{f}$, given the value of the frequency ($f = 2\text{s}^{-1}$).

Solution for (a)

Enter the known value into $T = \frac{1}{f}$

$$T = \frac{1}{2\ \text{s}^{-1}} = 0.\ 5\ \text{s}. \tag{13.6}$$

STRATEGY FOR (B)

Since one definition of wavelength is the distance a wave has traveled after one complete cycle—or one period—the values for the wavelength ($\lambda = 0.9\ \ m$) as well as the frequency are given. Therefore, we can use $v_{\text{w}} = f\lambda$ to find the wave velocity.

Solution for (b)

Enter the known values into $v_{\text{w}} = f\lambda$

$v_{\text{w}} = f\lambda = (2\ \text{s}^{-1})(0.9\ \text{m}) = 1.8\ \text{m/s}.$

Discussion

We could have also used the equation $v_{\text{w}} = \frac{\lambda}{T}$ to solve for the wave velocity since we already know the value of the period ($T = 0.5\ \text{s}$) from our calculation in part (a), and we would come up with the same answer.

Practice Problems

7. The frequency of a wave is 10 Hz. What is its period?
   a. The period of the wave is 100 s.
   b. The period of the wave is 10 s.
   c. The period of the wave is 0.01 s.
   d. The period of the wave is 0.1 s.

8. What is the velocity of a wave whose wavelength is 2 m and whose frequency is 5 Hz?
   a. 20 m/s
   b. 2.5 m/s
   c. 0.4 m/s
   d. 10 m/s

Check Your Understanding

9. What is the amplitude of a wave?
   a. A quarter of the total height of the wave
   b. Half of the total height of the wave
   c. Two times the total height of the wave
   d. Four times the total height of the wave

10. What is meant by the wavelength of a wave?
   a. The wavelength is the distance between adjacent identical parts of a wave, parallel to the direction of propagation.
   b. The wavelength is the distance between adjacent identical parts of a wave, perpendicular to the direction of propagation.
   c. The wavelength is the distance between a crest and the adjacent trough of a wave, parallel to the direction of propagation.
   d. The wavelength is the distance between a crest and the adjacent trough of a wave, perpendicular to the direction of propagation.

11. How can you mathematically express wave frequency in terms of wave period?
   a. $f = \frac{1}{T}$
   b. $f = \left(\frac{1}{T}\right)^2$
   c. $\text{f} = \text{T}$
   d. $f = (T)^2$

12. When is the wavelength directly proportional to the period of a wave?

a. When the velocity of the wave is halved
b. When the velocity of the wave is constant
c. When the velocity of the wave is doubled
d. When the velocity of the wave is tripled

13.3 Wave Interaction: Superposition and Interference

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Describe superposition of waves
- Describe interference of waves and distinguish between constructive and destructive interference of waves
- Describe the characteristics of standing waves
- Distinguish reflection from refraction of waves

Section Key Terms

antinode	constructive interference	destructive interference	inversion	nodes
reflection	refraction	standing wave	superposition	

Superposition of Waves

Most waves do not look very simple. They look more like the waves in Figure 13.10, rather than the simple water wave considered in the previous sections, which has a perfect sinusoidal shape.

Figure 13.10 These waves result from the superposition of several waves from different sources, producing a complex pattern. (Waterborough, Wikimedia Commons)

Most waves appear complex because they result from two or more simple waves that combine as they come together at the same place at the same time—a phenomenon called **superposition**.

Waves superimpose by adding their disturbances; each disturbance corresponds to a force, and all the forces add. If the disturbances are along the same line, then the resulting wave is a simple addition of the disturbances of the individual waves, that is, their amplitudes add.

Wave Interference

The two special cases of superposition that produce the simplest results are pure constructive interference and pure destructive interference.

Pure **constructive interference** occurs when two identical waves arrive at the same point exactly in phase. When waves are exactly in phase, the crests of the two waves are precisely aligned, as are the troughs. Refer to Figure 13.11. Because the disturbances add, the pure constructive interference of two waves with the same amplitude produces a wave that has twice the amplitude of the two individual waves, but has the same wavelength.

Figure 13.11 The pure constructive interference of two identical waves produces a wave with twice the amplitude but the same wavelength.

Figure 13.12 shows two identical waves that arrive exactly *out* of phase—that is, precisely aligned crest to trough—producing pure **destructive interference**. Because the disturbances are in opposite directions for this superposition, the resulting amplitude is zero for pure destructive interference; that is, the waves completely cancel out each other.

Figure 13.12 The pure destructive interference of two identical waves produces zero amplitude, or complete cancellation.

While pure constructive interference and pure destructive interference can occur, they are not very common because they require precisely aligned identical waves. The superposition of most waves that we see in nature produces a combination of constructive and destructive interferences.

Waves that are not results of pure constructive or destructive interference can vary from place to place and time to time. The sound from a stereo, for example, can be loud in one spot and soft in another. The varying loudness means that the sound waves add partially constructively and partially destructively at different locations. A stereo has at least two speakers that create sound waves, and waves can reflect from walls. All these waves superimpose.

An example of sounds that vary over time from constructive to destructive is found in the combined whine of jet engines heard by a stationary passenger. The volume of the combined sound can fluctuate up and down as the sound from the two engines varies in time from constructive to destructive.

The two previous examples considered waves that are similar—both stereo speakers generate sound waves with the same amplitude and wavelength, as do the jet engines. But what happens when two waves that are not similar, that is, having different amplitudes and wavelengths, are superimposed? An example of the superposition of two dissimilar waves is shown in Figure 13.13. Here again, the disturbances add and subtract, but they produce an even more complicated-looking wave. The resultant wave from the combined disturbances of two dissimilar waves looks much different than the idealized sinusoidal shape of a periodic wave.

Figure 13.13 The superposition of nonidentical waves exhibits both constructive and destructive interferences.

Virtual Physics

Wave Interference

Click to view content (http://www.openstax.org/l/28interference)

In this simulation, make waves with a dripping faucet, an audio speaker, or a laser by switching between the water, sound, and light tabs. Contrast and compare how the different types of waves behave. Try rotating the view from top to side to make observations. Then experiment with adding a second source or a pair of slits to create an interference pattern.

GRASP CHECK

In the water tab, compare the waves generated by one drip versus two drips. What happens to the amplitude of the waves when there are two drips? Is this constructive or destructive interference? Why would this be the case?

a. The amplitude of the water waves remains same because of the destructive interference as the drips of water hit the surface at the same time.
b. The amplitude of the water waves is canceled because of the destructive interference as the drips of water hit the surface at the same time.
c. The amplitude of water waves remains same because of the constructive interference as the drips of water hit the surface at the same time.
d. The amplitude of water waves doubles because of the constructive interference as the drips of water hit the surface at the same time.

Standing Waves

Sometimes waves do not seem to move and they appear to just stand in place, vibrating. Such waves are called **standing waves** and are formed by the superposition of two or more waves moving in opposite directions. The waves move through each other with their disturbances adding as they go by. If the two waves have the same amplitude and wavelength, then they alternate between constructive and destructive interference. Standing waves created by the superposition of two identical waves moving in opposite directions are illustrated in Figure 13.14.

Figure 13.14 A standing wave is created by the superposition of two identical waves moving in opposite directions. The oscillations are at fixed locations in space and result from alternating constructive and destructive interferences.

As an example, standing waves can be seen on the surface of a glass of milk in a refrigerator. The vibrations from the refrigerator motor create waves on the milk that oscillate up and down but do not seem to move across the surface. The two waves that

produce standing waves may be due to the reflections from the side of the glass.

Earthquakes can create standing waves and cause constructive and destructive interferences. As the earthquake waves travel along the surface of Earth and reflect off denser rocks, constructive interference occurs at certain points. As a result, areas closer to the epicenter are not damaged while areas farther from the epicenter are damaged.

Standing waves are also found on the strings of musical instruments and are due to reflections of waves from the ends of the string. Figure 13.15 and Figure 13.16 show three standing waves that can be created on a string that is fixed at both ends. When the wave reaches the fixed end, it has nowhere else to go but back where it came from, causing the reflection. The **nodes** are the points where the string does not move; more generally, the nodes are the points where the wave disturbance is zero in a standing wave. The fixed ends of strings must be nodes, too, because the string cannot move there.

The **antinode** is the location of maximum amplitude in standing waves. The standing waves on a string have a frequency that is related to the propagation speed v_w of the disturbance on the string. The wavelength λ is determined by the distance between the points where the string is fixed in place.

Figure 13.15 The figure shows a string oscillating with its maximum disturbance as the antinode.

Figure 13.16 The figure shows a string oscillating with multiple nodes.

Reflection and Refraction of Waves

As we saw in the case of standing waves on the strings of a musical instrument, **reflection** is the change in direction of a wave when it bounces off a barrier, such as a fixed end. When the wave hits the fixed end, it changes direction, returning to its source. As it is reflected, the wave experiences an **inversion**, which means that it flips vertically. If a wave hits the fixed end with a crest, it will return as a trough, and vice versa (Henderson 2015). Refer to Figure 13.17.

Figure 13.17 A wave is inverted after reflection from a fixed end.

TIPS FOR SUCCESS

If the end is not fixed, it is said to be a *free end*, and no inversion occurs. When the end is loosely attached, it reflects without

inversion, and when the end is not attached to anything, it does not reflect at all. You may have noticed this while changing the settings from Fixed End to Loose End to No End in the Waves on a String PhET simulation.

Rather than encountering a fixed end or barrier, waves sometimes pass from one medium into another, for instance, from air into water. Different types of media have different properties, such as density or depth, that affect how a wave travels through them. At the boundary between media, waves experience **refraction**—they change their path of propagation. As the wave bends, it also changes its speed and wavelength upon entering the new medium. Refer to Figure 13.18.

Figure 13.18 A wave refracts as it enters a different medium.

For example, water waves traveling from the deep end to the shallow end of a swimming pool experience refraction. They bend in a path closer to perpendicular to the surface of the water, propagate slower, and decrease in wavelength as they enter shallower water.

Check Your Understanding

13. What is the superposition of waves?
- a. When a single wave splits into two different waves at a point
- b. When two waves combine at the same place at the same time

14. How do waves superimpose on one another?
- a. By adding their frequencies
- b. By adding their wavelengths
- c. By adding their disturbances
- d. By adding their speeds

15. What is interference of waves?
- a. Interference is a superposition of two waves to form a resultant wave with higher or lower frequency.
- b. Interference is a superposition of two waves to form a wave of larger or smaller amplitude.
- c. Interference is a superposition of two waves to form a resultant wave with higher or lower velocity.
- d. Interference is a superposition of two waves to form a resultant wave with longer or shorter wavelength.

16. Is the following statement true or false? The two types of interference are constructive and destructive interferences.
- a. True
- b. False

17. What are standing waves?
- a. Waves that appear to remain in one place and do not seem to move
- b. Waves that seem to move along a trajectory

18. How are standing waves formed?
- a. Standing waves are formed by the superposition of two or more waves moving in opposite directions.
- b. Standing waves are formed by the superposition of two or more waves moving in the same direction.
- c. Standing waves are formed by the superposition of two or more waves moving in perpendicular directions.
- d. Standing waves are formed by the superposition of two or more waves moving in any arbitrary directions.

19. What is the reflection of a wave?
- a. The reflection of a wave is the change in amplitude of a wave when it bounces off a barrier.
- b. The reflection of a wave is the change in frequency of a wave when it bounces off a barrier.
- c. The reflection of a wave is the change in velocity of a wave when it bounces off a barrier.
- d. The reflection of a wave is the change in direction of a wave when it bounces off a barrier.

20. What is inversion of a wave?

a. Inversion occurs when a wave reflects off a fixed end, and the wave amplitude changes sign.
b. Inversion occurs when a wave reflects off a loose end, and the wave amplitude changes sign.
c. Inversion occurs when a wave reflects off a fixed end without the wave amplitude changing sign.
d. Inversion occurs when a wave reflects off a loose end without the wave amplitude changing sign.

KEY TERMS

antinode location of maximum amplitude in standing waves

constructive interference when two waves arrive at the same point exactly in phase; that is, the crests of the two waves are precisely aligned, as are the troughs

destructive interference when two identical waves arrive at the same point exactly out of phase that is precisely aligned crest to trough

inversion vertical flipping of a wave after reflection from a fixed end

longitudinal wave wave in which the disturbance is parallel to the direction of propagation

mechanical wave wave that requires a medium through which it can travel

medium solid, liquid, or gas material through which a wave propagates

nodes points where the string does not move; more generally, points where the wave disturbance is zero in a standing wave

periodic wave wave that repeats the same oscillation for several cycles and is associated with simple harmonic motion

pulse wave sudden disturbance with only one wave or a few waves generated

reflection change in direction of a wave at a boundary or fixed end

refraction bending of a wave as it passes from one medium to another medium with a different density

standing wave wave made by the superposition of two waves of the same amplitude and wavelength moving in opposite directions and which appears to vibrate in place

superposition phenomenon that occurs when two or more waves arrive at the same point

transverse wave wave in which the disturbance is perpendicular to the direction of propagation

wave disturbance that moves from its source and carries energy

wave velocity speed at which the disturbance moves; also called the propagation velocity or propagation speed

wavelength distance between adjacent identical parts of a wave

SECTION SUMMARY

13.1 Types of Waves

- A wave is a disturbance that moves from the point of creation and carries energy but not mass.
- Mechanical waves must travel through a medium.
- Sound waves, water waves, and earthquake waves are all examples of mechanical waves.
- Light is not a mechanical wave since it can travel through a vacuum.
- A periodic wave is a wave that repeats for several cycles, whereas a pulse wave has only one crest or a few crests and is associated with a sudden disturbance.
- Periodic waves are associated with simple harmonic motion.
- A transverse wave has a disturbance perpendicular to its direction of propagation, whereas a longitudinal wave has a disturbance parallel to its direction of propagation.

13.2 Wave Properties: Speed, Amplitude, Frequency, and Period

- A wave is a disturbance that moves from the point of creation at a wave velocity v_w.
- A wave has a wavelength λ, which is the distance between adjacent identical parts of the wave.
- The wave velocity and the wavelength are related to the wave's frequency and period by $v_w = \frac{\lambda}{T}$ or $v_w = f\lambda$.
- The time for one complete wave cycle is the period T.
- The number of waves per unit time is the frequency f.
- The wave frequency and the period are inversely related to one another.

13.3 Wave Interaction: Superposition and Interference

- Superposition is the combination of two waves at the same location.
- Constructive interference occurs when two identical waves are superimposed exactly in phase.
- Destructive interference occurs when two identical waves are superimposed exactly out of phase.
- A standing wave is a wave produced by the superposition of two waves. It varies in amplitude but does not propagate.
- The nodes are the points where there is no motion in standing waves.
- An antinode is the location of maximum amplitude of a standing wave.
- Reflection causes a wave to change direction.
- Inversion occurs when a wave reflects from a fixed end.
- Refraction causes a wave's path to bend and occurs when a wave passes from one medium into another medium with a different density.

KEY EQUATIONS

13.2 Wave Properties: Speed, Amplitude, Frequency, and Period

wave velocity $v_w = \frac{\lambda}{T}$ or $v_w = f\lambda$

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

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

CHAPTER REVIEW

Concept Items

13.1 Types of Waves

1. Do water waves push water from one place to another? Explain.
   a. No, water waves transfer only energy from one place to another.
   b. Yes, water waves transfer water from one place to another.

2. With reference to waves, what is a trough?
   a. the lowermost position of a wave
   b. the uppermost position of a wave
   c. the final position of a wave
   d. the initial position of the wave

3. Give an example of longitudinal waves.
   a. light waves
   b. water waves in a lake
   c. sound waves in air
   d. seismic waves in Earth's surface

4. What does the speed of a mechanical wave depend on?
   a. the properties of the material through which it travels
   b. the shape of the material through which it travels
   c. the size of the material through which it travels
   d. the color of the material through which it travels

13.2 Wave Properties: Speed, Amplitude, Frequency, and Period

5. Which characteristic of a transverse wave is measured along the direction of propagation?
   a. The amplitude of a transverse wave is measured along the direction of propagation.
   b. The amplitude and the wavelength of a transverse wave are measured along the direction of propagation.
   c. The wavelength of a transverse wave is measured along the direction of propagation.
   d. The displacement of the particles of the medium in a transverse wave is measured along the direction of propagation.

6. Which kind of seismic waves cannot travel through liquid?
   a. compressional waves
   b. P-waves
   c. longitudinal waves
   d. S-waves

7. What is the period of a wave?
   a. the time that a wave takes to complete a half cycle
   b. the time that a wave takes to complete one cycle
   c. the time that a wave takes to complete two cycles
   d. the time that a wave takes to complete four cycles

8. When the period of a wave increases, what happens to its frequency?
   a. Its frequency decreases.
   b. Its frequency increases.
   c. Its frequency remains the same.

13.3 Wave Interaction: Superposition and Interference

9. Is this statement true or false? The amplitudes of waves add up only if they are propagating in the same line.
   a. True
   b. False

10. Why is sound from a stereo louder in one part of the room and softer in another?
    a. Sound is louder in parts of the room where the density is greatest. Sound is softer in parts of the room where density is smallest.
    b. Sound is louder in parts of the room where the density is smallest. Sound is softer in parts of the room where density is greatest.
    c. Sound is louder in parts of the room where constructive interference occurs and softer in parts where destructive interference occurs.
    d. Sound is louder in parts of the room where destructive interference occurs and softer in parts where constructive interference occurs.

11. In standing waves on a string, what does the frequency depend on?
    a. The frequency depends on the propagation speed and the density of the string.

b. The frequency depends on the propagation speed and the length of the string.
c. The frequency depends on the density and the length of the string.
d. The frequency depends on the propagation speed, the density, and the length of the string.

12. Is the following statement true or false? Refraction is useful in fiber optic cables for transmitting signals.
a. False
b. True

13. What is refraction?
a. Refraction is the phenomenon in which waves change their path of propagation at the interface of two media with different densities.
b. Refraction is the phenomenon in which waves change their path of propagation at the interface of two media with the same density.
c. Refraction is the phenomenon in which waves become non-periodic at the boundary of two media with different densities.
d. Refraction is the phenomenon in which waves become non-periodic at the boundary of two media with the same density.

Critical Thinking Items

13.1 Types of Waves

14. Give an example of a wave that propagates only through a solid.
a. Light wave
b. Sound wave
c. Seismic wave
d. Surface wave

15. Can mechanical waves be periodic waves?
a. No, mechanical waves cannot be periodic waves.
b. Yes, mechanical waves can be periodic.

16. In a sound wave, which parameter of the medium varies with every cycle?
a. The density of the medium varies with every cycle.
b. The mass of the medium varies with every cycle.
c. The resistivity of the medium varies with every cycle.
d. The volume of the medium varies with every cycle.

17. What is a transverse wave in an earthquake called?
a. L-wave
b. P-wave
c. S-wave
d. R-wave

13.2 Wave Properties: Speed, Amplitude, Frequency, and Period

18. If the horizontal distance, that is, the distance in the direction of propagation, between a crest and the adjacent trough of a sine wave is 1 m, what is the wavelength of the wave?
a. 0.5 m
b. 1 m
c. 2 m
d. 4 m

19. How is the distance to the epicenter of an earthquake determined?
a. The wavelength difference between P-waves and S-waves is used to measure the distance to the epicenter.
b. The time difference between P-waves and S-waves is used to measure the distance to the epicenter.
c. The frequency difference between P-waves and S-waves is used to measure the distance to the epicenter.
d. The phase difference between P-waves and S-waves is used to measure the distance to the epicenter.

20. Two identical waves superimpose in pure constructive interference. What is the height of the resultant wave if the amplitude of each of the waves is 1 m?
a. 1 m
b. 2 m
c. 3 m
d. 4 m

13.3 Wave Interaction: Superposition and Interference

21. Two identical waves with an amplitude X superimpose in a way that pure constructive interference occurs. What is the amplitude of the resultant wave?
a. $\frac{X}{2}$
b. X
c. $2X$
d. X^2

22. In which kind of wave is the amplitude at each point constant?
a. Seismic waves
b. Pulse wave
c. Standing waves
d. Electromagnetic waves

23. Which property of a medium causes refraction?
a. Conductivity
b. Opacity
c. Ductility

d. Density

24. What is added together when two waves superimpose?
a. Amplitudes
b. Wavelengths
c. Velocities

Problems

13.2 Wave Properties: Speed, Amplitude, Frequency, and Period

25. If a seagull sitting in water bobs up and down once every 2 seconds and the distance between two crests of the water wave is 3 m, what is the velocity of the wave?
a. 1.5 m/s
b. 3 m/s
c. 6 m/s
d. 12 m/s

26. A boat in the trough of a wave takes 3 seconds to reach the highest point of the wave. The velocity of the wave is 5 m/s. What is its wavelength?
a. 0.83 m
b. 15 m
c. 30 m
d. 180 m

Performance Task

13.3 Wave Interaction: Superposition and Interference

27. Ocean waves repeatedly crash against beaches and coasts. Their energy can lead to erosion and collapse of land. Scientists and engineers need to study how waves interact with beaches in order to assess threats to coastal communities and construct breakwater systems. In this task, you will construct a wave tank and fill it with water. Simulate a beach by placing sand at one end. Create waves by moving a piece of wood or plastic up and down in the water. Measure or estimate the wavelength, period, frequency, and amplitude of the wave, and observe the effect of the wave on the sand. Produce waves of different amplitudes and frequencies, and record your observations each time. Use mathematical representations to demonstrate the relationships between different wave properties. Change the position of the sand to create a steeper beach and record your observations. Give a qualitative analysis of the effects of the waves on the beach. What kind of wave causes the most damage? At what height, wavelength, and frequency do waves *break*? How does the steepness of the beach affect the waves?

TEST PREP

Multiple Choice

13.1 Types of Waves

28. What kind of waves are sound waves?
a. Mechanical waves
b. Electromagnetic waves

29. What kind of a wave does a tuning fork create?
a. Pulse wave
b. Periodic wave
c. Electromagnetic wave

30. What kind of waves are electromagnetic waves?
a. Longitudinal waves
b. Transverse waves
c. Mechanical waves
d. P-waves

31. With reference to waves, what is a *disturbance?*
a. It refers to the resistance produced by some particles of a material.
b. It refers to an oscillation produced by some energy that creates a wave.
c. It refers to the wavelength of the wave.
d. It refers to the speed of the wave.

13.2 Wave Properties: Speed, Amplitude, Frequency, and Period

32. Which of these is not a characteristic of a wave?
a. amplitude
b. period
c. mass
d. velocity

33. If you are in a boat at a resting position, how much will your height change when you are hit by the peak of a wave with a height of 2 m?
a. 0 m
b. 1 m
c. 2 m
d. 4 m

34. What is the period of a wave with a frequency of 0.5 Hz?

a. 0.5 s
b. 1 s
c. 2 s
d. 3 s

35. What is the relation between the amplitude of a wave and its speed?
a. The amplitude of a wave is independent of its speed.
b. The amplitude of a wave is directly proportional to its speed.
c. The amplitude of a wave is directly proportional to the square of the inverse of its speed.
d. The amplitude of a wave is directly proportional to the inverse of its speed.

36. What does the speed of seismic waves depend on?
a. The speed of seismic waves depends on the size of the medium.
b. The speed of seismic waves depends on the shape of the medium.
c. The speed of seismic waves depends on the rigidity of the medium.

13.3 Wave Interaction: Superposition and Interference

37. What is added together when two waves superimpose?
a. amplitudes
b. wavelengths
c. velocities

38. Pure constructive interference occurs between two waves when they have the same _____.
a. frequency and are in phase
b. frequency and are out of phase
c. amplitude and are in phase
d. amplitude and are out of phase

39. What kind(s) of interference can occur between two identical waves moving in opposite directions?
a. Constructive interference only
b. Destructive interference only
c. Both constructive and destructive interference
d. Neither constructive nor destructive interference

40. What term refers to the bending of light at the junction of two media?
a. interference
b. diffraction
c. scattering
d. refraction

41. Which parameter of a wave gets affected after superposition?
a. wavelength
b. direction
c. amplitude
d. frequency

42. When do the amplitudes of two waves get added?
a. When their amplitudes are the same
b. When their amplitudes are different
c. When they propagate in perpendicular directions
d. When they are propagating along the same line in opposite directions

Short Answer

13.1 Types of Waves

43. Give an example of a non-mechanical wave.
a. A radio wave is an example of a nonmechanical wave.
b. A sound wave is an example of a nonmechanical wave.
c. A surface wave is an example of a nonmechanical wave.
d. A seismic wave is an example of a nonmechanical wave.

44. How is sound produced by an electronic speaker?
a. The cone of a speaker vibrates to create small changes in the temperature of the air.
b. The cone of a speaker vibrates to create small changes in the pressure of the air.
c. The cone of a speaker vibrates to create small changes in the volume of the air.
d. The cone of a speaker vibrates to create small changes in the resistance of the air.

45. What kind of wave is thunder?
a. Transverse wave
b. Pulse wave
c. R-wave
d. P-wave

46. Are all ocean waves perfectly sinusoidal?
a. No, all ocean waves are not perfectly sinusoidal.
b. Yes, all ocean waves are perfectly sinusoidal.

47. What are orbital progressive waves?
a. Waves that force the particles of the medium to follow a linear path from the crest to the trough
b. Waves that force the particles of the medium to follow a circular path from the crest to the trough
c. Waves that force the particles of the medium to follow a zigzag path from the crest to the trough
d. Waves that force the particles of the medium to

follow a random path from the crest to the trough

48. Give an example of orbital progressive waves.
a. light waves
b. ocean waves
c. sound waves
d. seismic waves

13.2 Wave Properties: Speed, Amplitude, Frequency, and Period

49. What is the relation between the amplitude and height of a transverse wave?
a. The height of a wave is half of its amplitude.
b. The height of a wave is equal to its amplitude.
c. The height of a wave is twice its amplitude.
d. The height of a wave is four times its amplitude.

50. If the amplitude of a water wave is 0.2 m and its frequency is 2 Hz, how much distance would a bird sitting on the water's surface move with every wave? How many times will it do this every second?
a. The bird will go up and down a distance of 0.4 m. It will do this twice per second.
b. The bird will go up and down a distance of 0.2 m. It will do this twice per second.
c. The bird will go up and down a distance of 0.4 m. It will do this once per second.
d. The bird will go up and down a distance of 0.2 m. It will do this once per second.

51. What is the relation between the amplitude and the frequency of a wave?
a. The amplitude and the frequency of a wave are independent of each other.
b. The amplitude and the frequency of a wave are equal.
c. The amplitude decreases with an increase in the frequency of a wave.
d. The amplitude increases with an increase in the frequency of a wave.

52. What is the relation between a wave's energy and its amplitude?
a. There is no relation between the energy and the amplitude of a wave.
b. The magnitude of the energy is equal to the magnitude of the amplitude of a wave.
c. The energy of a wave increases with an increase in the amplitude of the wave.
d. The energy of a wave decreases with an increase in the amplitude of a wave.

53. A wave travels 2 m every 2 cycles. What is its wavelength?
a. 4 m
b. 2 m
c. 0.5 m
d. 1 m

54. A water wave propagates in a river at 6 m/s. If the river moves in the opposite direction at 3 m/s, what is the effective velocity of the wave?
a. 3 m/s
b. 6 m/s
c. 9 m/s
d. 18 m/s

13.3 Wave Interaction: Superposition and Interference

55. Is this statement true or false? Spherical waves can superimpose.
a. True
b. False

56. Is this statement true or false? Waves can superimpose if their frequencies are different.
a. True
b. False

57. When does pure destructive interference occur?
a. When two waves with equal frequencies that are perfectly in phase and propagating along the same line superimpose.
b. When two waves with unequal frequencies that are perfectly in phase and propagating along the same line superimpose.
c. When two waves with unequal frequencies that are perfectly out of phase and propagating along the same line superimpose.
d. When two waves with equal frequencies that are perfectly out of phase and propagating along the same line superimpose.

58. Is this statement true or false? The amplitude of one wave is affected by the amplitude of another wave only when they are precisely aligned.
a. True
b. False

59. Why does a standing wave form on a guitar string?
a. due to superposition with the reflected waves from the ends of the string
b. due to superposition with the reflected waves from the walls of the room
c. due to superposition with waves generated from the body of the guitar

60. Is the following statement true or false? A standing wave is a superposition of two identical waves that are in phase and propagating in the same direction.

a. True
b. False

61. Why do water waves traveling from the deep end to the shallow end of a swimming pool experience refraction?
a. Because the pressure of water at the two ends of the pool is same
b. Because the pressures of water at the two ends of the pool are different
c. Because the density of water at the two ends of the pool is same
d. Because the density of water at the two ends of the pool is different

62. Is the statement true or false? Waves propagate faster in a less dense medium if the stiffness is the same.
a. True
b. False

Extended Response

13.1 Types of Waves

63. Why can light travel through outer space while sound cannot?
a. Sound waves are mechanical waves and require a medium to propagate. Light waves can travel through a vacuum.
b. Sound waves are electromagnetic waves and require a medium to propagate. Light waves can travel through a vacuum.
c. Light waves are mechanical waves and do not require a medium to propagate; sound waves require a medium to propagate.
d. Light waves are longitudinal waves and do not require a medium to propagate; sound waves require a medium to propagate.

64. Do periodic waves require a medium to travel through?
a. No, the requirement of a medium for propagation does not depend on whether the waves are pulse waves or periodic waves.
b. Yes, the requirement of a medium for propagation depends on whether the waves are pulse waves or periodic waves.

65. How is the propagation of sound in solids different from that in air?
a. Sound waves in solids are transverse, whereas in air, they are longitudinal.
b. Sound waves in solids are longitudinal, whereas in air, they are transverse.
c. Sound waves in solids can be both longitudinal and transverse, whereas in air, they are longitudinal.
d. Sound waves in solids are longitudinal, whereas in air, they can be both longitudinal and transverse.

13.2 Wave Properties: Speed, Amplitude, Frequency, and Period

66. A seagull is sitting in the water surface and a simple water wave passes under it. What sort of motion does the gull experience? Why?
a. The gull experiences mostly side-to-side motion and moves with the wave in its direction.
b. The gull experiences mostly side-to-side motion but does not move with the wave in its direction.
c. The gull experiences mostly up-and-down motion and moves with the wave in its direction.
d. The gull experiences mostly up-and-down motion but does not move in the direction of the wave.

67. Why does a good-quality speaker have a woofer and a tweeter?
a. In a good-quality speaker, sounds with high frequencies or short wavelengths are reproduced accurately by woofers, while sounds with low frequencies or long wavelengths are reproduced accurately by tweeters.
b. Sounds with high frequencies or short wavelengths are reproduced more accurately by tweeters, while sounds with low frequencies or long wavelengths are reproduced more accurately by woofers.

68. The time difference between a 2 km/s S-wave and a 6 km/s P-wave recorded at a certain point is 10 seconds. How far is the epicenter of the earthquake from that point?
a. 15 m
b. 30 m
c. 15 km
d. 30 km

13.3 Wave Interaction: Superposition and Interference

69. Why do water waves sometimes appear like a complex criss-cross pattern?
a. The crests and the troughs of waves traveling in the same direction combine to form a criss-cross pattern.
b. The crests and the troughs of waves traveling in different directions combine to form a criss-cross pattern.

70. What happens when two dissimilar waves interfere?
a. pure constructive interference
b. pure destructive interference
c. a combination of constructive and destructive

interference

71. Occasionally, during earthquakes, areas near the epicenter are not damaged while those farther away are damaged. Why could this occur?

a. Destructive interference results in waves with greater amplitudes being formed in places farther away from the epicenter.
b. Constructive interference results in waves with greater amplitudes being formed in places farther away from the epicenter.
c. The standing waves of great amplitudes are formed in places farther away from the epicenter.
d. The pulse waves of great amplitude are formed in places farther away from the epicenter.

72. Why does an object appear to be distorted when you view it through a glass of water?

a. The glass and the water reflect the light in different directions. Hence, the object appears to be distorted.
b. The glass and the water absorb the light by different amounts. Hence, the object appears to be distorted.
c. Water, air, and glass are media with different densities. Light rays refract and bend when they pass from one medium to another. Hence, the object appears to be distorted.
d. The glass and the water disperse the light into its components. Hence, the object appears to be distorted.

CHAPTER 14
Sound

Figure 14.1 This tree fell some time ago. When it fell, particles in the air were disturbed by the energy of the tree hitting the ground. This disturbance of matter, which our ears have evolved to detect, is called sound. (B.A. Bowen Photography)

Chapter Outline

14.1 Speed of Sound, Frequency, and Wavelength

14.2 Sound Intensity and Sound Level

14.3 Doppler Effect and Sonic Booms

14.4 Sound Interference and Resonance

INTRODUCTION If a tree falls in a forest (see Figure 14.1) and no one is there to hear it, does it make a sound? The answer to this old philosophical question depends on how you define sound. If sound only exists when someone is around to perceive it, then the falling tree produced no sound. However, in physics, we know that colliding objects can disturb the air, water or other matter surrounding them. As a result of the collision, the surrounding particles of matter began vibrating in a wave-like fashion. This is a sound wave. Consequently, if a tree collided with another object in space, no one would hear it, because no sound would be produced. This is because, in space, there is no air, water or other matter to be disturbed and produce sound waves. In this chapter, we'll learn more about the wave properties of sound, and explore hearing, as well as some special uses for sound.

14.1 Speed of Sound, Frequency, and Wavelength

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Relate the characteristics of waves to properties of sound waves
- Describe the speed of sound and how it changes in various media
- Relate the speed of sound to frequency and wavelength of a sound wave

Section Key Terms

rarefaction	sound

Properties of Sound Waves

Sound is a wave. More specifically, sound is defined to be a disturbance of matter that is transmitted from its source outward. A disturbance is anything that is moved from its state of equilibrium. Some sound waves can be characterized as periodic waves, which means that the atoms that make up the matter experience simple harmonic motion.

A vibrating string produces a sound wave as illustrated in Figure 14.2, Figure 14.3, and Figure 14.4. As the string oscillates back and forth, part of the string's energy goes into compressing and expanding the surrounding air. This creates slightly higher and lower pressures. The higher pressure... regions are compressions, and the low pressure regions are **rarefactions**. The pressure disturbance moves through the air as longitudinal waves with the same frequency as the string. Some of the energy is lost in the form of thermal energy transferred to the air. You may recall from the chapter on waves that areas of compression and rarefaction in longitudinal waves (such as sound) are analogous to crests and troughs in transverse waves.

Figure 14.2 A vibrating string moving to the right compresses the air in front of it and expands the air behind it.

Figure 14.3 As the string moves to the left, it creates another compression and rarefaction as the particles on the right move away from the string.

Figure 14.4 After many vibrations, there is a series of compressions and rarefactions that have been transmitted from the string as a sound wave. The graph shows gauge pressure (P_{gauge}) versus distance *x* from the source. Gauge pressure is the pressure relative to atmospheric pressure; it is positive for pressures above atmospheric pressure, and negative for pressures below it. For ordinary, everyday sounds, pressures vary only slightly from average atmospheric pressure.

The amplitude of a sound wave decreases with distance from its source, because the energy of the wave is spread over a larger and larger area. But some of the energy is also absorbed by objects, such as the eardrum in Figure 14.5, and some of the energy is converted to thermal energy in the air. Figure 14.4 shows a graph of gauge pressure versus distance from the vibrating string. From this figure, you can see that the compression of a longitudinal wave is analogous to the peak of a transverse wave, and the rarefaction of a longitudinal wave is analogous to the trough of a transverse wave. Just as a transverse wave alternates between peaks and troughs, a longitudinal wave alternates between compression and rarefaction.

Figure 14.5 Sound wave compressions and rarefactions travel up the ear canal and force the eardrum to vibrate. There is a net force on the eardrum, since the sound wave pressures differ from the atmospheric pressure found behind the eardrum. A complicated mechanism converts the vibrations to nerve impulses, which are then interpreted by the brain.

The Speed of Sound

The speed of sound varies greatly depending upon the medium it is traveling through. The speed of sound in a medium is determined by a combination of the medium's rigidity (or compressibility in gases) and its density. The more rigid (or less compressible) the medium, the faster the speed of sound. The greater the density of a medium, the slower the speed of sound. The speed of sound in air is low, because air is compressible. Because liquids and solids are relatively rigid and very difficult to compress, the speed of sound in such media is generally greater than in gases. Table 14.1 shows the speed of sound in various media. Since temperature affects density, the speed of sound varies with the temperature of the medium through which it's traveling to some extent, especially for gases.

Medium	v_w (m/s)
Gases at 0 °C	
Air	331
Carbon dioxide	259
Oxygen	316
Helium	965
Hydrogen	1290
Liquids at 20 °C	
Ethanol	1160
Mercury	1450
Water, fresh	1480
Sea water	1540
Human tissue	1540
Solids (longitudinal or bulk)	
Vulcanized rubber	54
Polyethylene	920
Marble	3810
Glass, Pyrex	5640
Lead	1960
Aluminum	5120
Steel	5960

Table 14.1 Speed of Sound in Various Media

The Relationship Between the Speed of Sound and the Frequency and Wavelength of a Sound Wave

Figure 14.6 When fireworks explode in the sky, the light energy is perceived before the sound energy. Sound travels more slowly than light does. (Dominic Alves, Flickr)

Sound, like all waves, travels at certain speeds through different media and has the properties of frequency and wavelength. Sound travels much slower than light—you can observe this while watching a fireworks display (see Figure 14.6), since the flash of an explosion is seen before its sound is heard.

The relationship between the speed of sound, its frequency, and wavelength is the same as for all waves:

$$v = f\lambda, \tag{14.1}$$

where v is the speed of sound (in units of m/s), f is its frequency (in units of hertz), and λ is its wavelength (in units of meters). Recall that wavelength is defined as the distance between adjacent identical parts of a wave. The wavelength of a sound, therefore, is the distance between adjacent identical parts of a sound wave. Just as the distance between adjacent crests in a transverse wave is one wavelength, the distance between adjacent compressions in a sound wave is also one wavelength, as shown in Figure 14.7. The frequency of a sound wave is the same as that of the source. For example, a tuning fork vibrating at a given frequency would produce sound waves that oscillate at that same frequency. The frequency of a sound is the number of waves that pass a point per unit time.

Figure 14.7 A sound wave emanates from a source vibrating at a frequency f, propagates at v, and has a wavelength λ .

One of the more important properties of sound is that its speed is nearly independent of frequency. If this were not the case, and high-frequency sounds traveled faster, for example, then the farther you were from a band in a football stadium, the more the sound from the low-pitch instruments would lag behind the high-pitch ones. But the music from all instruments arrives in cadence independent of distance, and so all frequencies must travel at nearly the same speed.

Recall that $v = f\lambda$, and in a given medium under fixed temperature and humidity, v is constant. Therefore, the relationship between f and λ is inverse: The higher the frequency, the shorter the wavelength of a sound wave.

The speed of sound can change when sound travels from one medium to another. However, the frequency usually remains the same because it is like a driven oscillation and maintains the frequency of the original source. If v changes and f remains the same, then the wavelength λ must change. Since $v = f\lambda$, the higher the speed of a sound, the greater its wavelength for a given frequency.

Virtual Physics

Sound

Click to view content (https://www.openstax.org/l/28sound)

This simulation lets you see sound waves. Adjust the frequency or amplitude (volume) and you can see and hear how the wave changes. Move the listener around and hear what she hears. Switch to the Two Source Interference tab or the Interference by Reflection tab to experiment with interference and reflection.

TIPS FOR SUCCESS

Make sure to have audio enabled and set to Listener rather than Speaker, or else the sound will not vary as you move the listener around.

GRASP CHECK

In the first tab, Listen to a Single Source, move the listener as far away from the speaker as possible, and then change the frequency of the sound wave. You may have noticed that there is a delay between the time when you change the setting and the time when you hear the sound get lower or higher in pitch. Why is this?

a. Because, intensity of the sound wave changes with the frequency.
b. Because, the speed of the sound wave changes when the frequency is changed.
c. Because, loudness of the sound wave takes time to adjust after a change in frequency.
d. Because it takes time for sound to reach the listener, so the listener perceives the new frequency of sound wave after a delay.

Is there a difference in the amount of delay depending on whether you make the frequency higher or lower? Why?

a. Yes, the speed of propagation depends only on the frequency of the wave.
b. Yes, the speed of propagation depends upon the wavelength of the wave, and wavelength changes as the frequency changes.
c. No, the speed of propagation depends only on the wavelength of the wave.
d. No, the speed of propagation is constant in a given medium; only the wavelength changes as the frequency changes.

Snap Lab

Voice as a Sound Wave

In this lab you will observe the effects of blowing and speaking into a piece of paper in order to compare and contrast different sound waves.

- sheet of paper
- tape
- table

Instructions

Procedure

1. Suspend a sheet of paper so that the top edge of the paper is fixed and the bottom edge is free to move. You could tape the top edge of the paper to the edge of a table, for example.
2. Gently blow air near the edge of the bottom of the sheet and note how the sheet moves.
3. Speak softly and then louder such that the sounds hit the edge of the bottom of the paper, and note how the sheet moves.
4. Interpret the results.

GRASP CHECK

Which sound wave property increases when you are speaking more loudly than softly?

a. amplitude of the wave
b. frequency of the wave
c. speed of the wave

d. wavelength of the wave

WORKED EXAMPLE

What Are the Wavelengths of Audible Sounds?

Calculate the wavelengths of sounds at the extremes of the audible range, 20 and 20,000 Hz, in conditions where sound travels at 348.7 m/s.

STRATEGY

To find wavelength from frequency, we can use $v = f\lambda$.

Solution

(1) Identify the knowns. The values for v and f are given.

(2) Solve the relationship between speed, frequency and wavelength for λ .

$$\lambda = \frac{v}{f}. \tag{14.2}$$

(3) Enter the speed and the minimum frequency to give the maximum wavelength.

$$\lambda_{\text{max}} = \frac{348.7 \text{ m/s}}{20 \text{ Hz}} = 17 \text{ m} \approx 20 \text{ m (1 sig. figure)} \tag{14.3}$$

(4) Enter the speed and the maximum frequency to give the minimum wavelength.

$$\lambda_{\text{min}} = \frac{348.7 \text{ m/s}}{20,000 \text{ Hz}} = 0.017 \text{ m} \approx 2 \text{ cm (1 sig. figure)} \tag{14.4}$$

Discussion

Because the product of f multiplied by λ equals a constant velocity in unchanging conditions, the smaller f is, the larger λ must be, and vice versa. Note that you can also easily rearrange the same formula to find frequency or velocity.

Practice Problems

1. What is the speed of a sound wave with frequency 2000 Hz and wavelength 0.4 m?
   a. 5×10^3 m/s
   b. 3.2×10^2 m/s
   c. 2×10^{-4} m/s
   d. 8×10^2 m/s

2. Dogs can hear frequencies of up to 45 kHz. What is the wavelength of a sound wave with this frequency traveling in air at 0°C?
   a. 2.0×10^7 m
   b. 1.5×10^7 m
   c. 1.4×10^2 m
   d. 7.4×10^{-3} m

LINKS TO PHYSICS

Echolocation

Figure 14.8 A bat uses sound echoes to find its way about and to catch prey. The time for the echo to return is directly proportional to the distance.

Echolocation is the use of reflected sound waves to locate and identify objects. It is used by animals such as bats, dolphins and whales, and is also imitated by humans in SONAR—Sound Navigation and Ranging—and echolocation technology.

Bats, dolphins and whales use echolocation to navigate and find food in their environment. They locate an object (or obstacle) by emitting a sound and then sensing the reflected sound waves. Since the speed of sound in air is constant, the time it takes for the sound to travel to the object and back gives the animal a sense of the distance between itself and the object. This is called *ranging*. Figure 14.8 shows a bat using echolocation to sense distances.

Echolocating animals identify an object by comparing the relative intensity of the sound waves returning to each ear to figure out the angle at which the sound waves were reflected. This gives information about the direction, size and shape of the object. Since there is a slight distance in position between the two ears of an animal, the sound may return to one of the ears with a bit of a delay, which also provides information about the position of the object. For example, if a bear is directly to the right of a bat, the echo will return to the bat's left ear later than to its right ear. If, however, the bear is directly ahead of the bat, the echo would return to both ears at the same time. For an animal without a sense of sight such as a bat, it is important to know *where* other animals are as well as *what* they are; their survival depends on it.

Principles of echolocation have been used to develop a variety of useful sensing technologies. SONAR, is used by submarines to detect objects underwater and measure water depth. Unlike animal echolocation, which relies on only one transmitter (a mouth) and two receivers (ears), manmade SONAR uses many transmitters and beams to get a more accurate reading of the environment. Radar technologies use the echo of radio waves to locate clouds and storm systems in weather forecasting, and to locate aircraft for air traffic control. Some new cars use echolocation technology to sense obstacles around the car, and warn the driver who may be about to hit something (or even to automatically parallel park). Echolocation technologies and training systems are being developed to help visually impaired people navigate their everyday environments.

GRASP CHECK

If a predator is directly to the left of a bat, how will the bat know?

a. The echo would return to the left ear first.
b. The echo would return to the right ear first.

Check Your Understanding

3. What is a rarefaction?
   a. Rarefaction is the high-pressure region created in a medium when a longitudinal wave passes through it.
   b. Rarefaction is the low-pressure region created in a medium when a longitudinal wave passes through it.
   c. Rarefaction is the highest point of amplitude of a sound wave.
   d. Rarefaction is the lowest point of amplitude of a sound wave.

4. What sort of motion do the particles of a medium experience when a sound wave passes through it?
   a. Simple harmonic motion

b. Circular motion
c. Random motion
d. Translational motion

5. What does the speed of sound depend on?
   a. The wavelength of the wave
   b. The size of the medium
   c. The frequency of the wave
   d. The properties of the medium

6. What property of a gas would affect the speed of sound traveling through it?
   a. The volume of the gas
   b. The flammability of the gas
   c. The mass of the gas
   d. The compressibility of the gas

14.2 Sound Intensity and Sound Level

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Relate amplitude of a wave to loudness and energy of a sound wave
- Describe the decibel scale for measuring sound intensity
- Solve problems involving the intensity of a sound wave
- Describe how humans produce and hear sounds

Section Key Terms

amplitude	decibel	hearing	loudness
pitch	sound intensity	sound intensity level	

Amplitude, Loudness and Energy of a Sound Wave

Figure 14.9 Noise on crowded roadways like this one in Delhi makes it hard to hear others unless they shout. (Lingaraj G J, Flickr)

In a quiet forest, you can sometimes hear a single leaf fall to the ground. But in a traffic jam filled with honking cars, you may have to shout just so the person next to you can hear Figure 14.9.The loudness of a sound is related to how energetically its source is vibrating. In cartoons showing a screaming person, the cartoonist often shows an open mouth with a vibrating uvula (the hanging tissue at the back of the mouth) to represent a loud sound coming from the throat. Figure 14.10 shows such a cartoon depiction of a bird loudly expressing its opinion.

A useful quantity for describing the loudness of sounds is called **sound intensity**. In general, the intensity of a wave is the power per unit area carried by the wave. Power is the rate at which energy is transferred by the wave. In equation form, intensity I is

$$I = \frac{P}{A}, \tag{14.5}$$

where P is the power through an area A. The SI unit for I is W/m^2. The intensity of a sound depends upon its pressure amplitude.

The relationship between the intensity of a sound wave and its pressure amplitude (or pressure variation Δp) is

$$I = \frac{(\Delta p)^2}{2\rho v_w}, \tag{14.6}$$

where ρ is the density of the material in which the sound wave travels, in units of kg/m^3, and v is the speed of sound in the medium, in units of m/s. Pressure amplitude has units of pascals (Pa) or N/m^2. Note that Δp is half the difference between the maximum and minimum pressure in the sound wave.

We can see from the equation that the intensity of a sound is proportional to its amplitude squared. The pressure variation is proportional to the amplitude of the oscillation, and so I varies as $(\Delta p)^2$. This relationship is consistent with the fact that the sound wave is produced by some vibration; the greater its pressure amplitude, the more the air is compressed during the vibration. Because the power of a sound wave is the rate at which energy is transferred, the energy of a sound wave is also proportional to its amplitude squared.

TIPS FOR SUCCESS

Pressure is usually denoted by capital P, but we are using a lowercase p for pressure in this case to distinguish it from power P above.

Figure 14.10 Graphs of the pressures in two sound waves of different intensities. The more intense sound is produced by a source that has larger-amplitude oscillations and has greater pressure maxima and minima. Because pressures are higher in the greater-intensity sound, it can exert larger forces on the objects it encounters.

The Decibel Scale

You may have noticed that when people talk about the loudness of a sound, they describe it in units of **decibels** rather than watts per meter squared. While sound intensity (in W/m^2) is the SI unit, the **sound intensity level** in decibels (dB) is more relevant for how humans perceive sounds. The way our ears perceive sound can be more accurately described by the logarithm of the intensity of a sound rather than the intensity of a sound directly. The sound intensity level β is defined to be

$$\beta\,(\text{dB}) = 10\log_{10}\left(\frac{I}{I_0}\right), \tag{14.7}$$

where I is sound intensity in watts per meter squared, and $I_0 = 10^{-12}\ W/m^2$ is a reference intensity. I_0 is chosen as the reference point because it is the lowest intensity of sound a person with normal hearing can perceive. The decibel level of a sound having an intensity of $10^{-12}\ W/m^2$ is $\beta = 0$ dB, because $\log_{10} 1 = 0$. That is, the threshold of human hearing is 0 decibels.

Each factor of 10 in intensity corresponds to 10 dB. For example, a 90 dB sound compared with a 60 dB sound is 30 dB greater, or three factors of 10 (that is, 10^3 times) as intense. Another example is that if one sound is 10^7 as intense as another, it is 70 dB higher.

Since β is defined in terms of a ratio, it is unit-less. The unit called *deci*bel (dB) is used to indicate that this ratio is multiplied by 10. The sound intensity level is not the same as sound intensity—it tells you the *level* of the sound relative to a reference intensity rather than the actual intensity.

Snap Lab

Feeling Sound

In this lab, you will play music with a heavy beat to literally feel the vibrations and explore what happens when the volume changes.

- CD player or portable electronic device connected to speakers
- rock or rap music CD or mp3
- a lightweight table

Procedure

1. Place the speakers on a light table, and start playing the CD or mp3.
2. Place your hand gently on the table next to the speakers.
3. Increase the volume and note the level when the table just begins to vibrate as the music plays.
4. Increase the reading on the volume control until it doubles. What has happened to the vibrations?

GRASP CHECK

Do you think that when you double the volume of a sound wave you are doubling the sound intensity level (in dB) or the sound intensity (in $\mathrm{W/m^2}$)? Why?

a. The sound intensity in $\mathrm{W/m^2}$, because it is a closer measure of how humans perceive sound.
b. The sound intensity level in dB because it is a closer measure of how humans perceive sound.
c. The sound intensity in $\mathrm{W/m^2}$ because it is the only unit to express the intensity of sound.
d. The sound intensity level in dB because it is the only unit to express the intensity of sound.

Solving Sound Wave Intensity Problems

WORKED EXAMPLE

Calculating Sound Intensity Levels: Sound Waves

Calculate the sound intensity level in decibels for a sound wave traveling in air at 0 °C and having a pressure amplitude of 0.656 Pa.

STRATEGY

We are given Δp, so we can calculate I using the equation $I = \frac{(\Delta p)^2}{2\rho v}$. Using I, we can calculate β straight from its definition in $\beta\,(d\mathrm{B}) = 10\log_{10}\left(\frac{I}{I_0}\right)\beta\,(\mathrm{dB}) = 10\log_{10}\left(\frac{I}{I_0}\right)$.

Solution

(1) Identify knowns:

Sound travels at 331 m/s in air at 0 °C.

Air has a density of 1.29 kg/m³ at atmospheric pressure and 0°C.

(2) Enter these values and the pressure amplitude into $I = \frac{(\Delta p)^2}{2\rho v_w}$.

$$I = \frac{(\Delta p)^2}{2\rho v_w} = \frac{(0.656\ \ \mathrm{Pa})^2}{2\left(1.29\ \mathrm{kg/m^3}\right)(331\ \ \mathrm{m/s})} = 5.04\times10^{-4}\ \mathrm{W/m^2}.$$

(3) Enter the value for I and the known value for I_0 into $\beta\,(\mathrm{dB}) = 10\log_{10}\left(\frac{I}{I_0}\right)$. Calculate to find the sound intensity level in decibels.

$$10\log_{10}\left(5.04\times10^{8}\right) = 10\,(8.70)\ \ \mathrm{dB} = 87.0\ \ \mathrm{dB}.$$

Discussion

This 87.0 dB sound has an intensity five times as great as an 80 dB sound. So a factor of five in intensity corresponds to a difference of 7 dB in sound intensity level. This value is true for any intensities differing by a factor of five.

WORKED EXAMPLE

Change Intensity Levels of a Sound: What Happens to the Decibel Level?

Show that if one sound is twice as intense as another, it has a sound level about 3 dB higher.

STRATEGY

You are given that the ratio of two intensities is 2 to 1, and are then asked to find the difference in their sound levels in decibels. You can solve this problem using of the properties of logarithms.

Solution

(1) Identify knowns:

The ratio of the two intensities is 2 to 1, or: $\frac{I_2}{I_1} = 2.00$.

We want to show that the difference in sound levels is about 3 dB. That is, we want to show

$$\beta_2 - \beta_1 = 3 \text{ dB}. \tag{14.8}$$

Note that

$$\log_{10} b - \log_{10} a = \log_{10}\left(\frac{b}{a}\right). \tag{14.9}$$

(2) Use the definition of β to get

$$\beta_2 - \beta_1 = 10 \log_{10}\left(\frac{I_2}{I_1}\right) = 10 \log_{10} 2.00 = 10\,(0.301) \text{ dB}. \tag{14.10}$$

Therefore, $\beta_2 - \beta_1 = 3.01 \text{ dB}$.

Discussion

This means that the two sound intensity levels differ by 3.01 dB, or about 3 dB, as advertised. Note that because only the ratio I_2/I_1 is given (and not the actual intensities), this result is true for any intensities that differ by a factor of two. For example, a 56.0 dB sound is twice as intense as a 53.0 dB sound, a 97.0 dB sound is half as intense as a 100 dB sound, and so on.

Practice Problems

7. Calculate the intensity of a wave if the power transferred is 10 W and the area through which the wave is transferred is 5 square meters.
   a. 200 W / m²
   b. 50 W / m²
   c. 0.5 W / m²
   d. 2 W / m²

8. Calculate the sound intensity for a sound wave traveling in air at 0°C and having a pressure amplitude of 0.90 Pa.
   a. $1.8 \times 10^{-3}\text{ W/m}^2$
   b. $4.2 \times 10^{-3}\text{ W/m}^2$
   c. $1.1 \times 10^{3}\text{ W/m}^2$
   d. $9.5 \times 10^{-4}\text{ W/m}^2$

Hearing and Voice

People create sounds by pushing air up through their lungs and through elastic folds in the throat called vocal cords. These folds

open and close rhythmically, creating a pressure buildup. As air travels up and past the vocal cords, it causes them to vibrate. This vibration escapes the mouth along with puffs of air as sound. A voice changes in pitch when the muscles of the larynx relax or tighten, changing the tension on the vocal chords. A voice becomes louder when air flow from the lungs increases, making the amplitude of the sound pressure wave greater.

Hearing is the perception of sound. It can give us plenty of information—such as pitch, loudness, and direction. Humans can normally hear frequencies ranging from approximately 20 to 20,000 Hz. Other animals have hearing ranges different from that of humans. Dogs can hear sounds as high as 45,000 Hz, whereas bats and dolphins can hear up to 110,000 Hz sounds. You may have noticed that dogs respond to the sound of a dog whistle which produces sound out of the range of human hearing.

Sounds below 20 Hz are called infrasound, whereas those above 20,000 Hz are ultrasound. The perception of frequency is called **pitch**, and the perception of intensity is called **loudness**.

The way we hear involves some interesting physics. The sound wave that hits our ear is a pressure wave. The ear converts sound waves into electrical nerve impulses, similar to a microphone.

Figure 14.11 shows the anatomy of the ear with its division into three parts: the outer ear or ear canal; the middle ear, which runs from the eardrum to the cochlea; and the inner ear, which is the cochlea itself. The body part normally referred to as the ear is technically called the pinna.

Figure 14.11 The illustration shows the anatomy of the human ear.

The outer ear, or ear canal, carries sound to the eardrum protected inside of the ear. The middle ear converts sound into mechanical vibrations and applies these vibrations to the cochlea. The lever system of the middle ear takes the force exerted on the eardrum by sound pressure variations, amplifies it and transmits it to the inner ear via the oval window. Two muscles in the middle ear protect the inner ear from very intense sounds. They react to intense sound in a few milliseconds and reduce the force transmitted to the cochlea. This protective reaction can also be triggered by your own voice, so that humming during a fireworks display, for example, can reduce noise damage.

Figure 14.12 shows the middle and inner ear in greater detail. As the middle ear bones vibrate, they vibrate the cochlea, which contains fluid. This creates pressure waves in the fluid that cause the tectorial membrane to vibrate. The motion of the tectorial membrane stimulates tiny cilia on specialized cells called hair cells. These hair cells, and their attached neurons, transform the motion of the tectorial membrane into electrical signals that are sent to the brain.

The tectorial membrane vibrates at different positions based on the frequency of the incoming sound. This allows us to detect the pitch of sound. Additional processing in the brain also allows us to determine which direction the sound is coming from (based on comparison of the sound's arrival time and intensity between our two ears).

Figure 14.12 The inner ear, or cochlea, is a coiled tube about 3 mm in diameter and 3 cm in length when uncoiled. As the stapes vibrates against the oval window, it creates pressure waves that travel through fluid in the cochlea. These waves vibrate the tectorial membrane, which bends the cilia and stimulates nerves in the organ of Corti. These nerves then send information about the sound to the brain.

FUN IN PHYSICS

Musical Instruments

Figure 14.13 Playing music, also known as "rocking out", involves creating vibrations using musical instruments. (John Norton)

Yet another way that people make sounds is through playing musical instruments (see the previous figure). Recall that the perception of frequency is called pitch. You may have noticed that the pitch range produced by an instrument tends to depend upon its size. Small instruments, such as a piccolo, typically make high-pitch sounds, while larger instruments, such as a tuba, typically make low-pitch sounds. High-pitch means small wavelength, and the size of a musical instrument is directly related to the wavelengths of sound it produces. So a small instrument creates short-wavelength sounds, just as a large instrument creates long-wavelength sounds.

Most of us have excellent relative pitch, which means that we can tell whether one sound has a different frequency from another. We can usually distinguish one sound from another if the frequencies of the two sounds differ by as little as 1 Hz. For example, 500.0 and 501.5 Hz are noticeably different.

Musical notes are particular sounds that can be produced by most instruments, and are the building blocks of a song. In Western music, musical notes have particular names, such as A-sharp, C, or E-flat. Some people can identify musical notes just by listening to them. This rare ability is called *perfect*, or *absolute, pitch*.

When a violin plays middle C, there is no mistaking it for a piano playing the same note. The reason is that each instrument produces a distinctive set of frequencies and intensities. We call our perception of these combinations of frequencies and intensities the *timbre* of the sound. It is more difficult to quantify timbre than loudness or pitch. Timbre is more subjective. Evocative adjectives such as dull, brilliant, warm, cold, pure, and rich are used to describe the timbre of a sound rather than

quantities with units, which makes for a difficult topic to dissect with physics. So the consideration of timbre takes us into the realm of perceptual psychology, where higher-level processes in the brain are dominant. This is also true for other perceptions of sound, such as music and noise. But as a teenager, you are likely already aware that one person's music may be another person's noise.

GRASP CHECK

If you turn up the volume of your stereo, will the pitch change? Why or why not?

a. No, because pitch does not depend on intensity.
b. Yes, because pitch is directly related to intensity.

Check Your Understanding

9. What is sound intensity?

a. Intensity is the energy per unit area carried by a wave.
b. Intensity is the energy per unit volume carried by a wave.
c. Intensity is the power per unit area carried by a wave.
d. Intensity is the power per unit volume carried by a wave.

10. How is power defined with reference to a sound wave?

a. Power is the rate at which energy is transferred by a sound wave.
b. Power is the rate at which mass is transferred by a sound wave.
c. Power is the rate at which amplitude of a sound wave changes.
d. Power is the rate at which wavelength of a sound wave changes.

11. What word or phrase is used to describe the loudness of sound?

a. frequency or oscillation
b. intensity level or decibel
c. timbre
d. pitch

12. What is the mathematical expression for sound intensity level β?

a. $\beta\,(\text{dB}) = 10\log_{10}\left(\frac{I_0}{I}\right)$
b. $\beta\,(\text{dB}) = 20\log_{10}\left(\frac{I}{I_0}\right)$
c. $\beta(dB) = 20\log_{10}\left(\frac{I_0}{I}\right)$
d. $\beta\,(\text{dB}) = 10\log_{10}\left(\frac{I}{I_0}\right)$

13. What is the range frequencies that humans are capable of hearing?

a. $20\text{ Hz to}200,000\text{ Hz}$
b. $2\text{ Hz to }50,000\text{ Hz}$
c. $2\text{ Hz to }2,000\text{ Hz}$
d. $20\text{ Hz to }20,000\text{ Hz}$

14. How do humans change the pitch of their voice?

a. Relaxing or tightening their glottis
b. Relaxing or tightening their uvula
c. Relaxing or tightening their tongue
d. Relaxing or tightening their larynx

References

Nave, R. Vocal sound production—HyperPhysics. Retrieved from http://hyperphysics.phy-astr.gsu.edu/hbase/music/voice.html

14.3 Doppler Effect and Sonic Booms

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Describe the Doppler effect of sound waves
- Explain a sonic boom
- Calculate the frequency shift of sound from a moving object by the Doppler shift formula, and calculate the speed of an object by the Doppler shift formula

Section Key Terms

Doppler effect	sonic boom

The Doppler Effect of Sound Waves

The **Doppler effect** is a change in the observed pitch of a sound, due to relative motion between the source and the observer. An example of the Doppler effect due to the motion of a source occurs when you are standing still, and the sound of a siren coming from an ambulance shifts from high-pitch to low-pitch as it passes by. The closer the ambulance is to you, the more sudden the shift. The faster the ambulance moves, the greater the shift. We also hear this shift in frequency for passing race cars, airplanes, and trains. An example of the Doppler effect with a stationary source and moving observer is if you ride a train past a stationary warning bell, you will hear the bell's frequency shift from high to low as you pass by.

What causes the Doppler effect? Let's compare three different scenarios: Sound waves emitted by a stationary source (Figure 14.14), sound waves emitted by a moving source (Figure 14.15), and sound waves emitted by a stationary source but heard by moving observers (Figure 14.16). In each case, the sound spreads out from the point where it was emitted.

If the source and observers are stationary, then observers on either side see the same wavelength and frequency as emitted by the source. But if the source is moving and continues to emit sound as it travels, then the air compressions (crests) become closer together in the direction in which it's traveling and farther apart in the direction it's traveling away from. Therefore, the wavelength is shorter in the direction the source is moving (on the right in Figure 14.15), and longer in the opposite direction (on the left in Figure 14.15).

Finally, if the observers move, as in Figure 14.16, the frequency at which they receive the compressions changes. The observer moving toward the source receives them at a higher frequency (and therefore shorter wavelength), and the person moving away from the source receives them at a lower frequency (and therefore longer wavelength).

Figure 14.14 Sounds emitted by a source spread out in spherical waves. Because the source, observers, and air are stationary, the wavelength and frequency are the same in all directions and to all observers.

Figure 14.15 Sounds emitted by a source moving to the right spread out from the points at which they were emitted. The wavelength is

reduced and, consequently, the frequency is increased in the direction of motion, so that the observer on the right hears a higher-pitch sound. The opposite is true for the observer on the left, where the wavelength is increased and the frequency is reduced.

Figure 14.16 The same effect is produced when the observers move relative to the source. Motion toward the source increases frequency as the observer on the right passes through more wave crests than she would if stationary. Motion away from the source decreases frequency as the observer on the left passes through fewer wave crests than he would if stationary.

We know that wavelength and frequency are related by $v = f\lambda$, where v is the fixed speed of sound. The sound moves in a medium and has the same speed v in that medium whether the source is moving or not. Therefore, f multiplied by λ is a constant. Because the observer on the right in Figure 14.15 receives a shorter wavelength, the frequency she perceives must be higher. Similarly, the observer on the left receives a longer wavelength and therefore perceives a lower frequency.

The same thing happens in Figure 14.16. A higher frequency is perceived by the observer moving toward the source, and a lower frequency is perceived by an observer moving away from the source. In general, then, relative motion of source and observer toward one another increases the perceived frequency. Relative motion apart decreases the perceived frequency. The greater the relative speed is, the greater the effect.

WATCH PHYSICS

Introduction to the Doppler Effect

This video explains the Doppler effect visually.

Click to view content (https://www.openstax.org/l/28doppler)

GRASP CHECK

If you are standing on the sidewalk facing the street and an ambulance drives by with its siren blaring, at what point will the frequency that you observe most closely match the actual frequency of the siren?

a. when it is coming toward you
b. when it is going away from you
c. when it is in front of you

For a stationary observer and a moving source of sound, the frequency (f_{obs}) of sound perceived by the observer is

$$f_{obs} = f_s \left(\frac{v_w}{v_w \pm v_s} \right), \tag{14.11}$$

where f_s is the frequency of sound from a source, v_s is the speed of the source along a line joining the source and observer, and v_w is the speed of sound. The minus sign is used for motion toward the observer and the plus sign for motion away from the observer.

TIPS FOR SUCCESS

Rather than just memorizing rules, which are easy to forget, it is better to think about the rules of an equation intuitively. Using a minus sign in $f_{obs} = f_s \left(\frac{v_w}{v_w \pm v_s} \right)$ will decrease the denominator and increase the observed frequency, which is consistent with the expected outcome of the Doppler effect when the source is moving toward the observer. Using a plus sign will increase the denominator and decrease the observed frequency, consistent with what you would expect for the source

moving away from the observer. This may be more helpful to keep in mind rather than memorizing the fact that "the minus sign is used for motion toward the observer and the plus sign for motion away from the observer."

Note that the greater the speed of the source, the greater the Doppler effect. Similarly, for a stationary source and moving observer, the frequency perceived by the observer f_{obs} is given by

$$f_{obs} = f_s \left(\frac{v_w \pm v_{obs}}{v_w} \right), \tag{14.12}$$

where v_{obs} is the speed of the observer along a line joining the source and observer. Here the plus sign is for motion toward the source, and the minus sign is for motion away from the source.

Sonic Booms

What happens to the sound produced by a moving source, such as a jet airplane, that approaches or even exceeds the speed of sound? Suppose a jet airplane is coming nearly straight at you, emitting a sound of frequency f_s. The greater the plane's speed, v_s, the greater the Doppler shift and the greater the value of f_{obs}. Now, as v_s approaches the speed of sound, v_w, f_{obs} approaches infinity, because the denominator in $f_{obs} = f_s \left(\frac{v_w}{v_w - v_s} \right)$ approaches zero.

This result means that at the speed of sound, in front of the source, each wave is superimposed on the previous one because the source moves forward at the speed of sound. The observer gets them all at the same instant, and so the frequency is theoretically infinite. If the source exceeds the speed of sound, no sound is received by the observer until the source has passed, so that the sounds from the source when it was approaching are stacked up with those from it when receding, creating a **sonic boom**. A sonic boom is a constructive interference of sound created by an object moving faster than sound.

An aircraft creates two sonic booms, one from its nose and one from its tail (see Figure 14.17). During television coverage of space shuttle landings, two distinct booms could often be heard. These were separated by exactly the time it would take the shuttle to pass by a point. Observers on the ground often do not observe the aircraft creating the sonic boom, because it has passed by before the shock wave reaches them. If the aircraft flies close by at low altitude, pressures in the sonic boom can be destructive enough to break windows. Because of this, supersonic flights are banned over populated areas of the United States.

Figure 14.17 Two sonic booms, created by the nose and tail of an aircraft, are observed on the ground after the plane has passed by.

Solving Problems Using the Doppler Shift Formula

WATCH PHYSICS

Doppler Effect Formula for Observed Frequency

This video explains the Doppler effect formula for cases when the source is moving toward the observer.

Click to view content (https://www.openstax.org/l/28dopplerform)

GRASP CHECK

Let's say that you have a rare phobia where you are afraid of the Doppler effect. If you see an ambulance coming your way, what would be the best strategy to minimize the Doppler effect and soothe your Doppleraphobia?

a. Stop moving and become stationary till it passes by.
b. Run toward the ambulance.
c. Run alongside the ambulance.

WATCH PHYSICS

Doppler Effect Formula When Source is Moving Away

This video explains the Doppler effect formula for cases when the source is moving away from the observer.

Click to view content (https://www.openstax.org/l/28doppleraway)

GRASP CHECK

Sal uses two different formulas for the Doppler effect-one for when the source is moving toward the observer and another for when the source is moving away. However, in this textbook we use only one formula. Explain.

a. The combined formula that can be used is, Use (+) when the source is moving toward the observer and (–) when the source is moving away from the observer.
b. The combined formula that can be used is, $f_{obs} = f_s\left(\frac{v_w \pm v_s}{v_w}\right)$. Use (+) when the source is moving away from the observer and (–) when the source is moving toward the observer.
c. The combined formula that can be used is, $f_{obs} = f_s\left(\frac{v_w}{v_w \pm v_s}\right)$. Use (+) when the source is moving toward the observer and (–) when the source is moving away from the observer.
d. The combined formula that can be used is, $f_{obs} = f_s\left(\frac{v_w}{v_w \pm v_s}\right)$. Use (+) when the source is moving away from the observer and (–) when the source is moving toward the observer.

WORKED EXAMPLE

Calculate Doppler Shift: A Train Horn

Suppose a train that has a 150 Hz horn is moving at 35 m/s in still air on a day when the speed of sound is 340 m/s. What frequencies are observed by a stationary person at the side of the tracks as the train approaches and after it passes?

Strategy

To find the observed frequency, $f_{obs} = f_s\left(\frac{v_w}{v_w \pm v_s}\right)$ must be used because the source is moving. The minus sign is used for the approaching train, and the plus sign for the receding train.

Solution

(1) Enter known values into $f_{obs} = f_s\left(\frac{v_w}{v_w - v_s}\right)$ to calculate the frequency observed by a stationary person as the train approaches:

$$f_{obs} = f_s\left(\frac{v_w}{v_w - v_s}\right) = (150\ \text{Hz})\left(\frac{340\ \text{m/s}}{340\ \text{m/s} - 35\ \text{m/s}}\right) = 167\ \text{Hz} \approx 170\ \text{Hz (2 sig. figs.)}$$

(2) Use the same equation but with the plus sign to find the frequency heard by a stationary person as the train recedes.

$$f_{obs} = f_s\left(\frac{v_w}{v_w + v_s}\right) = (150\ \text{Hz})\left(\frac{340\ \text{m/s}}{340\ \text{m/s} + 35\ \text{m/s}}\right) = 136\ \text{Hz} \approx 140\ \text{Hz (2 sig. figs.)}$$

Discussion

The numbers calculated are valid when the train is far enough away that the motion is nearly along the line joining the train and the observer. In both cases, the shift is significant and easily noticed. Note that the shift is approximately 20 Hz for motion toward and approximately 10 Hz for motion away. The shifts are not symmetric.

Practice Problems

15. What is the observed frequency when the source having frequency $3.0\,\text{kHz}$ is moving towards the observer at a speed of $1.0 \times 10^2\ \text{m/s}$ and the speed of sound is $331\ \text{m/s}$?
   a. $3.0\,\text{kHz}$
   b. $3.5\,\text{kHz}$
   c. $2.3\,\text{kHz}$
   d. $4.3\,\text{kHz}$

16. A train is moving away from you at a speed of $50.0\ \text{m/s}$. If you are standing still and hear the whistle at a frequency of $305\ \text{Hz}$, what is the actual frequency of the produced whistle? (Assume speed of sound to be $331\ \text{m/s}$.)
   a. $259\ \text{Hz}$
   b. $205\ \text{Hz}$
   c. $405\ \text{Hz}$
   d. $351\ \text{Hz}$

Check Your Understanding

17. What is the Doppler effect?
   a. The Doppler effect is a change in the observed speed of a sound due to the relative motion between the source and the observer.
   b. The Doppler effect is a change in the observed frequency of a sound due to the relative motion between the source and the observer.
   c. The Doppler effect is a change in the observed intensity of a sound due to the relative motion between the source and the observer.
   d. The Doppler effect is a change in the observed timbre of a sound, due to the relative motion between the source and the observer.

18. Give an example of the Doppler effect caused by motion of the source.
   a. The sound of a vehicle horn shifts from low-pitch to high-pitch as we move towards it.
   b. The sound of a vehicle horn shifts from low-pitch to high-pitch as we move away from it.
   c. The sound of a vehicle horn shifts from low-pitch to high-pitch as it passes by.
   d. The sound of a vehicle horn shifts from high-pitch to low-pitch as it passes by.

19. What is a sonic boom?
   a. It is a destructive interference of sound created by an object moving faster than sound.
   b. It is a constructive interference of sound created by an object moving faster than sound.
   c. It is a destructive interference of sound created by an object moving slower than sound.
   d. It is a constructive interference of sound created by an object moving slower than sound.

20. What is the relation between speed of source and value of observed frequency when the source is moving towards the observer?
   a. They are independent of each other.
   b. The greater the speed, the greater the value of observed frequency.
   c. The greater the speed, the smaller the value of observed frequency.
   d. The speed of the sound is directly proportional to the square of the frequency observed.

14.4 Sound Interference and Resonance

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Describe resonance and beats
- Define fundamental frequency and harmonic series
- Contrast an open-pipe and closed-pipe resonator
- Solve problems involving harmonic series and beat frequency

Section Key Terms

beat	beat frequency	damping	fundamental	harmonics
natural frequency	overtones	resonance	resonate	

Resonance and Beats

Sit in front of a piano sometime and sing a loud brief note at it while pushing down on the sustain pedal. It will sing the same note back at you—the strings that have the same frequencies as your voice, are resonating in response to the forces from the sound waves that you sent to them. This is a good example of the fact that objects—in this case, piano strings—can be forced to oscillate but oscillate best at their natural frequency.

A driving force (such as your voice in the example) puts energy into a system at a certain frequency, which is not necessarily the same as the natural frequency of the system. Over time the energy dissipates, and the amplitude gradually reduces to zero- this is called **damping**. The **natural frequency** is the frequency at which a system would oscillate if there were no driving and no damping force. The phenomenon of driving a system with a frequency equal to its natural frequency is called **resonance**, and a system being driven at its natural frequency is said to **resonate**.

Most of us have played with toys where an object bobs up and down on an elastic band, something like the paddle ball suspended from a finger in Figure 14.18. At first you hold your finger steady, and the ball bounces up and down with a small amount of damping. If you move your finger up and down slowly, the ball will follow along without bouncing much on its own. As you increase the frequency at which you move your finger up and down, the ball will respond by oscillating with increasing amplitude. When you drive the ball at its natural frequency, the ball's oscillations increase in amplitude with each oscillation for as long as you drive it. As the driving frequency gets progressively higher than the resonant or natural frequency, the amplitude of the oscillations becomes smaller, until the oscillations nearly disappear and your finger simply moves up and down with little effect on the ball.

Figure 14.18 The paddle ball on its rubber band moves in response to the finger supporting it. If the finger moves with the natural frequency of the ball on the rubber band, then a resonance is achieved, and the amplitude of the ball's oscillations increases dramatically. At higher and lower driving frequencies, energy is transferred to the ball less efficiently, and it responds with lower-amplitude oscillations.

Another example is that when you tune a radio, you adjust its resonant frequency so that it oscillates only at the desired station's broadcast (driving) frequency. Also, a child on a swing is driven (pushed) by a parent at the swing's natural frequency to reach the maximum amplitude (height). In all of these cases, the efficiency of energy transfer from the driving force into the oscillator is best at resonance.

Figure 14.19 Some types of headphones use the phenomena of constructive and destructive interference to cancel out outside noises.

All sound resonances are due to constructive and destructive interference. Only the resonant frequencies interfere constructively to form standing waves, while others interfere destructively and are absent. From the toot made by blowing over a bottle to the recognizability of a great singer's voice, resonance and standing waves play a vital role in sound.

Interference happens to all types of waves, including sound waves. In fact, one way to support that something *is a wave* is to observe interference effects. Figure 14.19 shows a set of headphones that employs a clever use of sound interference to cancel noise. To get destructive interference, a fast electronic analysis is performed, and a second sound is introduced with its maxima and minima exactly reversed from the incoming noise.

In addition to resonance, superposition of waves can also create beats. **Beats** are produced by the superposition of two waves with slightly different frequencies but the same amplitude. The waves alternate in time between constructive interference and destructive interference, giving the resultant wave an amplitude that varies over time. (See the resultant wave in Figure 14.20).

This wave fluctuates in amplitude, or beats, with a frequency called the **beat frequency**. The equation for beat frequency is

$$f_B = |f_1 - f_2|, \tag{14.13}$$

where f_1 and f_2 are the frequencies of the two original waves. If the two frequencies of sound waves are similar, then what we hear is an average frequency that gets louder and softer at the beat frequency.

TIPS FOR SUCCESS

Don't confuse the beat frequency with the regular frequency of a wave resulting from superposition. While the beat frequency is given by the formula above, and describes the frequency of the beats, the actual frequency of the wave resulting from superposition is the average of the frequencies of the two original waves.

Figure 14.20 Beats are produced by the superposition of two waves of slightly different frequencies but identical amplitudes. The waves alternate in time between constructive interference and destructive interference, giving the resulting wave a time-varying amplitude.

Virtual Physics

Wave Interference

Click to view content (https://www.openstax.org/l/28interference)

For this activity, switch to the Sound tab. Turn on the Sound option, and experiment with changing the frequency and amplitude, and adding in a second speaker and a barrier.

GRASP CHECK

According to the graph, what happens to the amplitude of pressure over time. What is this phenomenon called, and what causes it ?

a. The amplitude decreases over time. This phenomenon is called damping. It is caused by the dissipation of energy.
b. The amplitude increases over time. This phenomenon is called feedback. It is caused by the gathering of energy.
c. The amplitude oscillates over time. This phenomenon is called echoing. It is caused by fluctuations in energy.

Fundamental Frequency and Harmonics

Suppose we hold a tuning fork near the end of a tube that is closed at the other end, as shown in Figure 14.21, Figure 14.22, and Figure 14.23. If the tuning fork has just the right frequency, the air column in the tube resonates loudly, but at most frequencies it vibrates very little. This means that the air column has only certain natural frequencies. The figures show how a resonance at the lowest of these natural frequencies is formed. A disturbance travels down the tube at the speed of sound and bounces off the closed end. If the tube is just the right length, the reflected sound arrives back at the tuning fork exactly half a cycle later, and it interferes constructively with the continuing sound produced by the tuning fork. The incoming and reflected sounds form a standing wave in the tube as shown.

Figure 14.21 Resonance of air in a tube closed at one end, caused by a tuning fork. A disturbance moves down the tube.

Figure 14.22 Resonance of air in a tube closed at one end, caused by a tuning fork. The disturbance reflects from the closed end of the tube.

Figure 14.23 Resonance of air in a tube closed at one end, caused by a tuning fork. If the length of the tube *L* is just right, the disturbance gets back to the tuning fork half a cycle later and interferes constructively with the continuing sound from the tuning fork. This interference forms a standing wave, and the air column resonates.

The standing wave formed in the tube has its maximum air displacement (an antinode) at the open end, and no displacement (a

node) at the closed end. Recall from the last chapter on waves that motion is unconstrained at the antinode, and halted at the node. The distance from a node to an antinode is one-fourth of a wavelength, and this equals the length of the tube; therefore, $\lambda = 4L$. This same resonance can be produced by a vibration introduced at or near the closed end of the tube, as shown in Figure 14.24.

Figure 14.24 The same standing wave is created in the tube by a vibration introduced near its closed end.

Since maximum air displacements are possible at the open end and none at the closed end, there are other, shorter wavelengths that can resonate in the tube see Figure 14.25). Here the standing wave has three-fourths of its wavelength in the tube, or $L = (3/4)\lambda'$, so that $\lambda' = 4L/3$. There is a whole series of shorter-wavelength and higher-frequency sounds that resonate in the tube.

We use specific terms for the resonances in any system. The lowest resonant frequency is called the **fundamental**, while all higher resonant frequencies are called **overtones**. All resonant frequencies are multiples of the fundamental, and are called **harmonics**. The fundamental is the first harmonic, the first overtone is the second harmonic, and so on. Figure 14.26 shows the fundamental and the first three overtones (the first four harmonics) in a tube closed at one end.

Figure 14.25 Another resonance for a tube closed at one end. This has maximum air displacements at the open end, and none at the closed end. The wavelength is shorter, with three-fourths λ' equaling the length of the tube, so that $\lambda' = 4L/3$. This higher-frequency vibration is the first overtone.

Figure 14.26 The fundamental and three lowest overtones for a tube closed at one end. All have maximum air displacements at the open end and none at the closed end.

The fundamental and overtones can be present at the same time in a variety of combinations. For example, the note middle C on a trumpet sounds very different from middle C on a clarinet, even though both instruments are basically modified versions of a

tube closed at one end. The fundamental frequency is the same (and usually the most intense), but the overtones and their mix of intensities are different. This mix is what gives musical instruments (and human voices) their distinctive characteristics, whether they have air columns, strings, or drumheads. In fact, much of our speech is determined by shaping the cavity formed by the throat and mouth and positioning the tongue to adjust the fundamental and combination of overtones.

Open-Pipe and Closed-Pipe Resonators

The resonant frequencies of a tube closed at one end (known as a closed-pipe resonator) are $f_n = n\frac{v}{4L},\ n = 1, 3, 5...,$

where f_1 is the fundamental, f_3 is the first overtone, and so on. Note that the resonant frequencies depend on the speed of sound v and on the length of the tube L.

Another type of tube is one that is *open* at both ends (known as an open-pipe resonator). Examples are some organ pipes, flutes, and oboes. The air columns in tubes open at both ends have maximum air displacements at both ends. (See Figure 14.27). Standing waves form as shown.

Figure 14.27 The resonant frequencies of a tube open at both ends are shown, including the fundamental and the first three overtones. In all cases the maximum air displacements occur at both ends of the tube, giving it different natural frequencies than a tube closed at one end.

The resonant frequencies of an open-pipe resonator are

$$f_n = n\frac{v}{2L},\ n = 1, 2, 3...,$$

where f_1 is the fundamental, f_2 is the first overtone, f_3 is the second overtone, and so on. Note that a tube open at both ends has a fundamental frequency twice what it would have if closed at one end. It also has a different spectrum of overtones than a tube closed at one end. So if you had two tubes with the same fundamental frequency but one was open at both ends and the other was closed at one end, they would sound different when played because they have different overtones.

Middle C, for example, would sound richer played on an open tube since it has more overtones. An open-pipe resonator has more overtones than a closed-pipe resonator because it has even multiples of the fundamental as well as odd, whereas a closed tube has only odd multiples.

In this section we have covered resonance and standing waves for wind instruments, but vibrating strings on stringed instruments also resonate and have fundamentals and overtones similar to those for wind instruments.

Solving Problems Involving Harmonic Series and Beat Frequency

WORKED EXAMPLE

Finding the Length of a Tube for a Closed-Pipe Resonator

If sound travels through the air at a speed of 344 m/s, what should be the length of a tube closed at one end to have a fundamental frequency of 128 Hz?

Strategy

The length L can be found by rearranging the equation $f_n = n\frac{v}{4L}$.

Solution

(1) Identify knowns.

- The fundamental frequency is 128 Hz.
- The speed of sound is 344 m/s.

(2) Use $f_n = n\frac{v_w}{4L}$ to find the fundamental frequency ($n = 1$).

$$f_1 = \frac{v}{4L} \tag{14.14}$$

(3) Solve this equation for length.

$$L = \frac{v}{4f_1} \tag{14.15}$$

(4) Enter the values of the speed of sound and frequency into the expression for L.

$$L = \frac{v}{4f_1} = \frac{344\ \text{m/s}}{4\,(128\ \text{Hz})} = 0.672\ \text{m} \tag{14.16}$$

Discussion

Many wind instruments are modified tubes that have finger holes, valves, and other devices for changing the length of the resonating air column and therefore, the frequency of the note played. Horns producing very low frequencies, such as tubas, require tubes so long that they are coiled into loops.

WORKED EXAMPLE

Finding the Third Overtone in an Open-Pipe Resonator

If a tube that's open at both ends has a fundamental frequency of 120 Hz, what is the frequency of its third overtone?

Strategy

Since we already know the value of the fundamental frequency (n = 1), we can solve for the third overtone (n = 4) using the equation $f_n = n\frac{v}{2L}$.

Solution

Since fundamental frequency (n = 1) is

$$f_1 = \frac{v}{2L}, \tag{14.17}$$

and

$$f_4 = 4\frac{v}{2L}, f_4 = 4f_1 = 4(120\ \text{Hz}) = 480\ \text{Hz}. \tag{14.18}$$

Discussion

To solve this problem, it wasn't necessary to know the length of the tube or the speed of the air because of the relationship between the fundamental and the third overtone. This example was of an open-pipe resonator; note that for a closed-pipe resonator, the third overtone has a value of n = 7 (not n = 4).

WORKED EXAMPLE

Using Beat Frequency to Tune a Piano

Piano tuners use beats routinely in their work. When comparing a note with a tuning fork, they listen for beats and adjust the string until the beats go away (to zero frequency). If a piano tuner hears two beats per second, and the tuning fork has a

frequency of 256 Hz, what are the possible frequencies of the piano?

Strategy

Since we already know that the beat frequency f_B is 2, and one of the frequencies (let's say f_2) is 256 Hz, we can use the equation $f_B = |f_1 - f_2|$ to solve for the frequency of the piano f_1.

Solution

Since $f_B = |f_1 - f_2|$,

we know that either $f_B = f_1 - f_2$ or $-f_B = f_1 - f_2$.

Solving for f_1,

$$f_1 = f_B + f_2 \text{ or } f_1 = -f_B + f_2. \tag{14.19}$$

Substituting in values,

$$f_1 = 2 + 256 \text{ Hz or } f_1 = -2 + 256 \text{ Hz} \tag{14.20}$$

So,

$$f_1 = 258 \text{ Hz or } 254 \text{ Hz}. \tag{14.21}$$

Discussion

The piano tuner might not initially be able to tell simply by listening whether the frequency of the piano is too high or too low and must tune it by trial and error, making an adjustment and then testing it again. If there are even more beats after the adjustment, then the tuner knows that he went in the wrong direction.

Practice Problems

21. Two sound waves have frequencies 250 Hz and 280 Hz. What is the beat frequency produced by their superposition?

a. 290 Hz
b. 265 Hz
c. 60 Hz
d. 30 Hz

22. What is the length of a pipe closed at one end with fundamental frequency 350 Hz? (Assume the speed of sound in air is 331 m/s.)

a. 26 cm
b. 26 m
c. 24 m
d. 24 cm

Check Your Understanding

23. What is damping?

a. Over time the energy increases and the amplitude gradually reduces to zero. This is called damping.
b. Over time the energy dissipates and the amplitude gradually increases. This is called damping.
c. Over time the energy increases and the amplitude gradually increases. This is called damping.
d. Over time the energy dissipates and the amplitude gradually reduces to zero. This is called damping.

24. What is resonance? When can you say that the system is resonating?

a. The phenomenon of driving a system with a frequency equal to its natural frequency is called resonance, and a system being driven at its natural frequency is said to resonate.
b. The phenomenon of driving a system with a frequency higher than its natural frequency is called resonance, and a system being driven at its natural frequency does not resonate.
c. The phenomenon of driving a system with a frequency equal to its natural frequency is called resonance, and a system being driven at its natural frequency does not resonate.
d. The phenomenon of driving a system with a frequency higher than its natural frequency is called resonance, and a system being driven at its natural frequency is said to resonate.

25. In the tuning fork and tube experiment, in case a standing wave is formed, at what point on the tube is the maximum disturbance from the tuning fork observed? Recall that the tube has one open end and one closed end.
   a. At the midpoint of the tube
   b. Both ends of the tube
   c. At the closed end of the tube
   d. At the open end of the tube

26. In the tuning fork and tube experiment, when will the air column produce the loudest sound?
   a. If the tuning fork vibrates at a frequency twice that of the natural frequency of the air column.
   b. If the tuning fork vibrates at a frequency lower than the natural frequency of the air column.
   c. If the tuning fork vibrates at a frequency higher than the natural frequency of the air column.
   d. If the tuning fork vibrates at a frequency equal to the natural frequency of the air column.

27. What is a closed-pipe resonator?
   a. A pipe or cylindrical air column closed at both ends
   b. A pipe with an antinode at the closed end
   c. A pipe with a node at the open end
   d. A pipe or cylindrical air column closed at one end

28. Give two examples of open-pipe resonators.
   a. piano, violin
   b. drum, tabla
   c. rlectric guitar, acoustic guitar
   d. flute, oboe

KEY TERMS

amplitude the amount that matter is disrupted during a sound wave, as measured by the difference in height between the crests and troughs of the sound wave.

beat a phenomenon produced by the superposition of two waves with slightly different frequencies but the same amplitude

beat frequency the frequency of the amplitude fluctuations of a wave

damping the reduction in amplitude over time as the energy of an oscillation dissipates

decibel a unit used to describe sound intensity levels

Doppler effect an alteration in the observed frequency of a sound due to relative motion between the source and the observer

fundamental the lowest-frequency resonance

harmonics the term used to refer to the fundamental and its overtones

hearing the perception of sound

loudness the perception of sound intensity

natural frequency the frequency at which a system would oscillate if there were no driving and no damping forces

overtones all resonant frequencies higher than the fundamental

pitch the perception of the frequency of a sound

rarefaction a low-pressure region in a sound wave

resonance the phenomenon of driving a system with a frequency equal to the system's natural frequency

resonate to drive a system at its natural frequency

sonic boom a constructive interference of sound created by an object moving faster than sound

sound a disturbance of matter that is transmitted from its source outward by longitudinal waves

sound intensity the power per unit area carried by a sound wave

sound intensity level the level of sound relative to a fixed standard related to human hearing

SECTION SUMMARY

14.1 Speed of Sound, Frequency, and Wavelength

- Sound is one type of wave.
- Sound is a disturbance of matter that is transmitted from its source outward in the form of longitudinal waves.
- The relationship of the speed of sound v, its frequency f, and its wavelength λ is given by $v = f\lambda$, which is the same relationship given for all waves.
- The speed of sound depends upon the medium through which the sound wave is travelling.
- In a given medium at a specific temperature (or density), the speed of sound v is the same for all frequencies and wavelengths.

14.2 Sound Intensity and Sound Level

- The intensity of a sound is proportional to its amplitude squared.
- The energy of a sound wave is also proportional to its amplitude squared.
- Sound intensity level in decibels (dB) is more relevant for how humans perceive sounds than sound intensity (in W/m^2), even though sound intensity is the SI unit.
- Sound intensity level is not the same as sound intensity—it tells you the *level* of the sound relative to a reference intensity rather than the actual intensity.
- Hearing is the perception of sound and involves that transformation of sound waves into vibrations of parts within the ear. These vibrations are then transformed into neural signals that are interpreted by the brain.
- People create sounds by pushing air up through their lungs and through elastic folds in the throat called vocal cords.

14.3 Doppler Effect and Sonic Booms

- The Doppler effect is a shift in the observed frequency of a sound due to motion of either the source or the observer.
- The observed frequency is greater than the actual source's frequency when the source and the observer are moving closer together, either by the source moving toward the observer or the observer moving toward the source.
- A sonic boom is constructive interference of sound created by an object moving faster than sound.

14.4 Sound Interference and Resonance

- A system's natural frequency is the frequency at which the system will oscillate if not affected by driving or damping forces.
- A periodic force driving a harmonic oscillator at its natural frequency produces resonance. The system is said to resonate.
- Beats occur when waves of slightly different frequencies are superimposed.
- In air columns, the lowest-frequency resonance is called the fundamental, whereas all higher resonant frequencies are called overtones. Collectively, they are

called harmonics.

- The resonant frequencies of a tube closed at one end are $f_n = n\frac{v}{4L},\ n = 1, 3, 5...$, where f_1 is the fundamental and L is the length of the tube.
- The resonant frequencies of a tube open at both ends are $f_n = n\frac{v}{2L},\ n = 1, 2, 3...$

KEY EQUATIONS

14.1 Speed of Sound, Frequency, and Wavelength

speed of sound	$v = f\lambda$

14.2 Sound Intensity and Sound Level

intensity	$I = \frac{P}{A}$
sound intensity	$I = \frac{(\Delta p)^2}{2\rho v_w}$
sound intensity level	$\beta\,(\text{dB}) = 10\log_{10}\left(\frac{I}{I_0}\right)$

14.3 Doppler Effect and Sonic Booms

Doppler effect observed frequency (moving source)	$f_{obs} = f_s\left(\frac{v_w}{v_w \pm v_s}\right)$
Doppler effect observed frequency (moving observer)	$f_{obs} = f_s\left(\frac{v_w \pm v_{obs}}{v_w}\right)$

14.4 Sound Interference and Resonance

beat frequency	$f_B = \lvert f_1 - f_2 \rvert$
resonant frequencies of a closed-pipe resonator	$f_n = n\frac{v}{4L},\ n = 1, 3, 5...$
resonant frequencies of an open-pipe resonator	$f_n = n\frac{v}{2L},\ n = 1, 2, 3...$

CHAPTER REVIEW

Concept Items

14.1 Speed of Sound, Frequency, and Wavelength

1. What is the amplitude of a sound wave perceived by the human ear?
   a. loudness
   b. pitch
   c. intensity
   d. timbre

2. The compressibility of air and hydrogen is almost the same. Which factor is the reason that sound travels faster in hydrogen than in air?
   a. Hydrogen is more dense than air.
   b. Hydrogen is less dense than air.
   c. Hydrogen atoms are heavier than air molecules.
   d. Hydrogen atoms are lighter than air molecules.

14.2 Sound Intensity and Sound Level

3. What is the mathematical relationship between intensity, power, and area?
   a. $I = \frac{P}{A^2}$
   b. $I = PA$
   c. $I = \frac{A}{P}$
   d. $I = \frac{P}{A}$

4. How does the "decibel" get its name?
   a. The meaning of deci is “hundred” and the number of decibels is one-hundredth of the logarithm to base 10 of the ratio of two sound intensities.
   b. The meaning of deci is "ten" and the number of decibels is one-tenth of the logarithm to base 10 of the ratio of two sound intensities.
   c. The meaning of deci is “one-hundredth” and the number of decibels is hundred times the logarithm to base 10 of the ratio of two sound intensities.
   d. The meaning of deci is “one-tenth” and the number of decibels is ten times the logarithm to base 10 of the ratio of two sound intensities.

5. What is “timbre” of sound?
   a. Timbre is the quality of the sound that distinguishes it from other sound

b. Timbre is the loudness of the sound that distinguishes it from other sound.
c. Timbre is the pitch of the sound that distinguishes it from other sound.
d. Timbre is the wavelength of the sound that distinguishes it from other sound.

14.3 Doppler Effect and Sonic Booms

6. Two sources of sound producing the same frequency are moving towards you at different speeds. Which one would sound more high-pitched?
a. the one moving slower
b. the one moving faster

7. When the speed of the source matches the speed of sound, what happens to the amplitude of the sound wave? Why?
a. It approaches zero. This is because all wave crests are superimposed on one another through constructive interference.
b. It approaches infinity. This is because all wave crests are superimposed on one another through constructive interference.
c. It approaches zero, because all wave crests are superimposed on one another through destructive interference.
d. It approaches infinity, because all wave crests are superimposed on one another through destructive interference.

8. What is the mathematical expression for the frequency perceived by the observer in the case of a stationary observer and a moving source?
a. $f_{obs} = f_s \left(\frac{v_w}{v_s \pm v_w}\right)$
b. $f_{obs} = f_s \left(\frac{v_w \pm v_s}{v_w}\right)$
c. $f_{obs} = f_s \left(\frac{v_s \pm v_w}{v_w}\right)$
d. $f_{obs} = f_s \left(\frac{v_w}{v_w \pm v_s}\right)$

14.4 Sound Interference and Resonance

9. When does a yo-yo travel the farthest from the finger?
a. when the amplitude of the finger moving up and down is greater than the amplitude of the yo-yo
b. when the amplitude of the finger moving up and down is less than the amplitude of the yo-yo
c. when the frequency of the finger moving up and down is equal to the resonant frequency of the yo-yo
d. when the frequency of the finger moving up and down is different from the resonant frequency of the yo-yo

10. What is the difference between harmonics and overtones?
a. Harmonics are all multiples of the fundamental frequency. The first overtone is actually the first harmonic.
b. Harmonics are all multiples of the fundamental frequency. The first overtone is actually the second harmonic.
c. Harmonics are all multiples of the fundamental frequency. The second overtone is actually the first harmonic.
d. Harmonics are all multiples of the fundamental frequency. The third overtone is actually the second harmonic.

11. What kind of waves form in pipe resonators?
a. damped waves
b. propagating waves
c. high-frequency waves
d. standing waves

12. What is the natural frequency of a system?
a. The natural frequency is the frequency at which a system oscillates when it undergoes forced vibration.
b. The natural frequency is the frequency at which a system oscillates when it undergoes damped oscillation.
c. The natural frequency is the frequency at which a system oscillates when it undergoes free vibration without a driving force or damping.
d. The natural frequency is the frequency at which a system oscillates when it undergoes forced vibration with damping.

Critical Thinking Items

14.1 Speed of Sound, Frequency, and Wavelength

13. What can be said about the frequency of a monotonous sound?
a. It decreases with time.
b. It decreases with distance.
c. It increases with distance.
d. It remains constant.

14. A scientist notices that a sound travels faster through a solid material than through the air. Which of the following can explain this?
a. Solid materials are denser than air.
b. Solid materials are less dense than air.
c. A solid is more rigid than air.
d. A solid is easier to compress than air.

14.2 Sound Intensity and Sound Level

15. Which property of the wave is related to its intensity? How?
   a. The frequency of the wave is related to the intensity of the sound. The larger-frequency oscillations indicate greater pressure maxima and minima, and the pressure is higher in greater-intensity sound.
   b. The wavelength of the wave is related to the intensity of the sound. The longer-wavelength oscillations indicate greater pressure maxima and minima, and the pressure is higher in greater-intensity sound.
   c. The amplitude of the wave is related to the intensity of the sound. The larger-amplitude oscillations indicate greater pressure maxima and minima, and the pressure is higher in greater-intensity sound.
   d. The speed of the wave is related to the intensity of the sound. The higher-speed oscillations indicate greater pressure maxima and minima, and the pressure is higher in greater-intensity sound.

16. Why is decibel (dB) used to describe loudness of sound?
   a. Because, human ears have an inverse response to the amplitude of sound.
   b. Because, human ears have an inverse response to the intensity of sound.
   c. Because, the way our ears perceive sound can be more accurately described by the amplitude of a sound rather than the intensity of a sound directly.
   d. Because, the way our ears perceive sound can be more accurately described by the logarithm of the intensity of a sound rather than the intensity of a sound directly.

17. How can humming while shooting a gun reduce ear damage?
   a. Humming can trigger those two muscles in the outer ear that react to intense sound produced while shooting and reduce the force transmitted to the cochlea.
   b. Humming can trigger those three muscles in the outer ear that react to intense sound produced while shooting and reduce the force transmitted to the cochlea.
   c. Humming can trigger those two muscles in the middle ear that react to intense sound produced while shooting and reduce the force transmitted to the cochlea.
   d. Humming can trigger those three muscles in the middle ear that react to intense sound produced while shooting and reduce the force transmitted to the cochlea.

18. A particular sound, S1, has an intensity 3 times that of another sound, S2. What is the difference in sound intensity levels measured in decibels?
   a. 9.54 dB
   b. 6.02 dB
   c. 3.01 dB
   d. 4.77 dB

14.3 Doppler Effect and Sonic Booms

19. When the source of sound is moving through the air, does the speed of sound change with respect to a stationary person standing nearby?
   a. Yes
   b. No

20. Why is no sound heard by the observer when an object approaches him at a speed faster than that of sound?
   a. If the source exceeds the speed of sound, then destructive interference occurs and no sound is heard by the observer when an object approaches him.
   b. If the source exceeds the speed of sound, the frequency of sound produced is beyond the audible range of sound.
   c. If the source exceeds the speed of sound, all the sound waves produced approach minimum intensity and no sound is heard by the observer when an object approaches him.
   d. If the source exceeds the speed of sound, all the sound waves produced are behind the source. Hence, the observer hears the sound only after the source has passed.

21. Does the Doppler effect occur when the source and observer are both moving towards each other? If so, how would this affect the perceived frequency?
   a. Yes, the perceived frequency will be even lower in this case than if only one of the two were moving.
   b. No, the Doppler effect occurs only when an observer is moving towards a source.
   c. No, the Doppler effect occurs only when a source is moving towards an observer.
   d. Yes, the perceived frequency will be even higher in this case than if only one of the two were moving.

14.4 Sound Interference and Resonance

22. When does the amplitude of an oscillating system become maximum?
   a. When two sound waves interfere destructively.
   b. When the driving force produces a transverse wave in the system.
   c. When the driving force of the oscillator to the oscillating system is at a maximum amplitude.
   d. When the frequency of the oscillator equals the

natural frequency of the oscillating system.

23. How can a standing wave be formed with the help of a tuning fork and a closed-end tube of appropriate length?
   a. If the tube is just the right length, the reflected sound arrives back at the tuning fork exactly half a cycle later, and it interferes constructively with the continuing sound produced by the tuning fork.
   b. If the tube is just the right length, the reflected sound arrives back at the tuning fork exactly half a cycle later, and it interferes destructively with the continuing sound produced by the tuning fork.
   c. If the tube is just the right length, the reflected sound arrives back at the tuning fork exactly one full cycle later, and it interferes constructively with the continuing sound produced by the tuning fork.
   d. If the tube is just the right length, the reflected sound arrives back at the tuning fork exactly one full cycle later, and it interferes destructively with the continuing sound produced by the tuning fork.

24. A tube open at both ends has a fundamental frequency of 500 Hz. What will the frequency be if one end is closed?
   a. 1000 Hz
   b. 500 Hz
   c. 125 Hz
   d. 250 Hz

Problems

14.1 Speed of Sound, Frequency, and Wavelength

25. A bat produces a sound at 17, 250 Hz and wavelength 0.019 m. What is the speed of the sound?
   a. 1.7×10^6 m/s
   b. 8.6×10^5 m/s
   c. 1.15×10^{-6} m/s
   d. 3.28×10^2 m/s

26. A sound wave with frequency of 80 Hz is traveling through air at 0°C. By how much will its wavelength change when it enters aluminum?
   a. 68 m
   b. 64 m
   c. 4 m
   d. 60 m

14.2 Sound Intensity and Sound Level

27. Calculate the sound intensity for a sound wave traveling through air at 15° C and having a pressure amplitude of 0.80 Pa. (Hint—Speed of sound in air at 15° C is 340 m/s .)
   a. 9.6×10^{-3} W / m^2
   b. 7.7×10^{-3} W / m^2
   c. 9.6×10^{-4} W / m^2
   d. 7.7×10^{-4} W / m^2

28. The sound level in dB of a sound traveling through air at 0°C is 97 dB. Calculate its pressure amplitude.
   a. 4.3 Pa
   b. 0.20 Pa
   c. 0.04 Pa
   d. 2.1 Pa

14.3 Doppler Effect and Sonic Booms

29. An ambulance is moving away from you. You are standing still and you hear its siren at a frequency of 101 Hz. You know that the actual frequency of the siren is 105 Hz. What is the speed of the ambulance? (Assume the speed of sound to be 331 m/s.)
   a. 17.07 m/s
   b. 16.55 m/s
   c. 14.59 m/s
   d. 13.1 m/s

30. An ambulance passes you at a speed of 15.0 m/s. If its siren has a frequency of 995 Hz, what is difference in the frequencies you perceive before and after it passes you? (Assume the speed of sound in air is 331 m/s.)
   a. 47.0 Hz
   b. 43.0 Hz
   c. 94.9 Hz
   d. 90.0 Hz

14.4 Sound Interference and Resonance

31. What is the length of an open-pipe resonator with a fundamental frequency of 400.0 Hz? (Assume the speed of sound is 331 m/s.)
   a. 165.1 cm
   b. 82.22 cm
   c. 20.25 cm
   d. 41.38 cm

32. An open-pipe resonator has a fundamental frequency of 250 Hz. By how much would its length have to be changed to get a fundamental frequency of 300.0 Hz? (Assume the speed of sound is 331 m/s.)
   a. 77.32 cm
   b. 44.09 cm
   c. 32.16 cm
   d. 11.03 cm

Performance Task

14.4 Sound Interference and Resonance

33. Design and make an open air resonator capable of playing at least three different pitches (frequencies) of sound using a selection of bamboo of varying widths and lengths, which can be obtained at a local hardware store. Choose a piece of bamboo for creating a musical pipe. Calculate the length required for a certain frequency to resonate and then mark the locations where holes should be placed in the pipe to achieve their desired pitches. Use a simple hand drill or ask your wood shop department for help drilling holes. Use tuning forks to test and calibrate your instrument. Demonstrate your pipe for the class.

TEST PREP

Multiple Choice

14.1 Speed of Sound, Frequency, and Wavelength

34. What properties does a loud, shrill whistle have?
- a. high amplitude, high frequency
- b. high amplitude, low frequency
- c. low amplitude, high frequency
- d. low amplitude, low frequency

35. What is the speed of sound in fresh water at 20 degrees Celsius?
- a. 5960 m/s
- b. 1540 m/s
- c. 331 m/s
- d. 1480 m/s

36. A tuning fork oscillates at a frequency of 512 Hz, creating sound waves. How many waves will reach the eardrum of a person near that fork in 2 seconds?
- a. 512
- b. 128
- c. 256
- d. 1024

37. Why does the amplitude of a sound wave decrease with distance from its source?
- a. The amplitude of a sound wave decreases with distance from its source, because the frequency of the sound wave decreases.
- b. The amplitude of a sound wave decreases with distance from its source, because the speed of the sound wave decreases.
- c. The amplitude of a sound wave decreases with distance from its source, because the wavelength of the sound wave increases.
- d. The amplitude of a sound wave decreases with distance from its source, because the energy of the wave is spread over a larger and larger area.

38. Does the elasticity of the medium affect the speed of sound? How?
- a. No, there is no relationship that exists between the speed of sound and elasticity of the medium.
- b. Yes. When particles are more easily compressed in a medium, sound does not travel as quickly through the medium.
- c. Yes. When the particles in a medium do not compress much, sound does not travel as quickly through the medium.
- d. No, the elasticity of a medium affects frequency and wavelength, not wave speed.

14.2 Sound Intensity and Sound Level

39. Which of the following terms is a useful quantity to describe the loudness of a sound?
- a. intensity
- b. frequency
- c. pitch
- d. wavelength

40. What is the unit of sound intensity level?
- a. decibels
- b. hertz
- c. watts

41. If a particular sound S1 is 5 times more intense than another sound S2, then what is the difference in sound intensity levels in dB for these two sounds?
- a. 5 dB
- b. 6 dB
- c. 7 dB

42. By what minimum amount should frequencies vary for humans to be able to distinguish two separate sounds?
- a. 100 Hz
- b. 10 Hz
- c. 5 Hz
- d. 1 Hz

43. Why is I^o chosen as the reference for sound intensity?
- a. Because, it is the highest intensity of sound a person with normal hearing can perceive at a frequency of 100 Hz.
- b. Because, it is the lowest intensity of sound a person with normal hearing can perceive at a frequency of 100 Hz.

c. Because, it is the highest intensity of sound a person with normal hearing can perceive at a frequency of 1000 Hz.
d. Because, it is the lowest intensity of sound a person with normal hearing can perceive at a frequency of 1000 Hz.

14.3 Doppler Effect and Sonic Booms

44. In which of the following situations is the Doppler effect absent?
a. The source and the observer are moving towards each other.
b. The observer is moving toward the source.
c. The source is moving away from the observer.
d. Neither the source nor the observer is moving relative to one another.

45. What does the occurrence of the sonic boom depend on?
a. speed of the source
b. frequency of source
c. amplitude of source
d. distance of observer from the source

46. What is the observed frequency when the observer is moving away from the source at 125 m/s? The source frequency is 237 Hz and the speed of sound is 325 m/s .
a. 303 Hz
b. 259 Hz
c. 201 Hz
d. 146 Hz

47. How will your perceived frequency change if the source is moving towards you?
a. The frequency will become lower.
b. The frequency will become higher.

14.4 Sound Interference and Resonance

48. Observation of which phenomenon can be considered proof that something is a wave?
a. interference
b. noise
c. reflection
d. conduction

49. Which of the resonant frequencies has the greatest amplitude?
a. The first harmonic
b. The second harmonic
c. The first overtone
d. The second overtone

50. What is the fundamental frequency of an open-pipe resonator?
a. 3 v/2 L
b. 2 v/L
c. v/L
d. v/2 L

51. What is the beat frequency produced by the superposition of two waves with frequencies 300 Hz and 340 Hz?
a. 640 Hz
b. 320 Hz
c. 20 Hz
d. 40 Hz

Short Answer

14.1 Speed of Sound, Frequency, and Wavelength

52. What component of a longitudinal sound wave is analogous to a trough of a transverse wave?
a. compression
b. rarefaction
c. node
d. antinode

53. What is the frequency of a sound wave as perceived by the human ear?
a. timbre
b. loudness
c. intensity
d. pitch

54. What properties of a solid determine the speed of sound traveling through it?
a. mass and density
b. rigidity and density
c. volume and density
d. shape and rigidity

55. Does the density of a medium affect the speed of sound?
a. No
b. Yes

56. Does a bat make use of the properties of sound waves to locate its prey?
a. No
b. Yes

57. Do the properties of a sound wave change when it travels from one medium to another?
a. No
b. Yes

14.2 Sound Intensity and Sound Level

58. When a passing driver has his stereo turned up, you cannot even hear what the person next to you is saying. Why is this so?

a. The sound from the passing car's stereo has a higher amplitude and hence higher intensity compared to the intensity of the sound coming from the person next to you. The higher intensity corresponds to greater loudness, so the first sound dominates the second.

b. The sound from the passing car's stereo has a higher amplitude and hence lower intensity compared to the intensity of the sound coming from the person next to you. The lower intensity corresponds to greater loudness, so the first sound dominates the second.

c. The sound from the passing car's stereo has a higher frequency and hence higher intensity compared to the intensity of the sound coming from the person next to you. The higher frequency corresponds to greater loudness so the first sound dominates the second.

d. The sound from the passing car's stereo has a lower frequency and hence higher intensity compared to the intensity of the sound coming from the person next to you. The lower frequency corresponds to greater loudness, so the first sound dominates the second.

59. For a constant area, what is the relationship between intensity of a sound wave and power?

a. The intensity is inversely proportional to the power transmitted by the wave, for a constant area.

b. The intensity is inversely proportional to the square of the power transmitted by the wave, for a constant area. $I = \frac{1}{P^2}$

c. The intensity is directly proportional to the square of the power transmitted by the wave, for a constant area. $I = \mathrm{P}^2$

d. The intensity is directly proportional to the power transmitted by the wave, for a constant area. $I = \frac{P}{A}$

60. What does I stand for in the equation $\beta(\mathrm{dB}) = 10\ \log_{10}\left(\frac{I}{I_0}\right)$? What is its unit?

a. Yes, I is the sound intensity in watts per meter squared in the equation, $\beta(\mathrm{dB}) = 10\log_{10}\left(\frac{I}{I_0}\right)$.

b. I is the sound illuminance and its SI unit is lumen per meter squared.

c. I is the sound intensity and its SI unit is watts per meter cubed.

d. I is the sound intensity and its SI unit is watts per meter squared.

61. Why is the reference intensity $I_0 = 10^{-12}\ \mathrm{W/m^2}$?

a. The upper limit of human hearing is 100 decibels, i.e. $\beta = 100\ \mathrm{dB}$. For $\beta = 100\ \mathrm{dB}$, $I_0 = 10^{-12}\ \mathrm{W/m^2}$.

b. The lower threshold of human hearing is 10 decibels, i.e. $\beta = 10\ \mathrm{dB}$. For $\beta = 10\ \mathrm{dB}$, $I_0 = 10^{-12}\ \mathrm{W/m^2}$

c. The upper limit of human hearing is 10 decibels, i.e. $\beta = 10\ \mathrm{dB}$. For $\beta = 10\ \mathrm{dB}$, $I_0 = 10^{-12}\ \mathrm{W/m^2}$

d. The lower threshold of human hearing is 0 decibels, i.e., $\beta = 0\ \mathrm{dB}$. For $\beta = 0\ \mathrm{dB}$, $I_0 = 10^{-12}\ \mathrm{W/m^2}$

62. Given that the sound intensity level of a particular wave is 82 dB, what will be the sound intensity for that wave?

a. $I = 1.6 \times 10^{-6}\ \mathrm{W/m^2}$

b. $I = 82 \times 10^{-12}\ \mathrm{W/m^2}$

c. $I = 8.2 \times 10^{-12}\ \mathrm{W/m^2}$

d. $I = 1.6 \times 10^{-4}\ \mathrm{W/m^2}$

63. For a sound wave with intensity $1.58 \times 10^{-4}\ \mathrm{W/m^2}$, calculate the pressure amplitude given that the sound travels through air at $0^\circ\mathrm{C}$.

a. 0.734 Pa

b. 3.67 Pa

c. 0.135 Pa

d. 0.367 Pa

64. Which nerve carries auditory information to the brain?

a. buccal nerve

b. peroneal nerve

c. cochlear nerve

d. mandibular nerve

65. Why do some smaller instruments, such as piccolos, produce higher-pitched sounds than larger instruments, such as tubas?

a. Smaller instruments produce sounds with shorter wavelengths, and thus higher frequencies.

b. Smaller instruments produce longer wavelength, and thus higher amplitude, sounds.

c. Smaller instruments produce lower amplitude, and thus longer wavelength sounds.

d. Smaller instruments produce higher amplitude, and thus lower frequency, sounds.

14.3 Doppler Effect and Sonic Booms

66. How will your perceived frequency change if you move away from a stationary source of sound?

a. The frequency will become lower.

b. The frequency will be doubled.

c. The frequency will be tripled.

d. The frequency will become higher.

67. True or false—The Doppler effect also occurs with waves other than sound waves.
a. False
b. True

68. A source of sound is moving towards you. How will what you hear change if the speed of the source increases?
a. The sound will become more high-pitched.
b. The sound will become more low-pitched.
c. The pitch of the sound will not change.

69. Do sonic booms continue to be created when an object is traveling at supersonic speeds?
a. No, a sonic boom is created only when the source exceeds the speed of sound.
b. Yes, sonic booms continue to be created when an object is traveling at supersonic speeds.

70. Suppose you are driving at a speed of 20.0 m/s and you hear the sound of a bell at a frequency of 400.0 Hz. What is the actual frequency of the bell if the speed of sound is 335 m/s?
a. $f_s = 401\text{ Hz or } f_s = 315\text{ Hz}$
b. $f_s = 385\text{ Hz or } f_s = 419\text{ Hz}$
c. $f_s = 415\text{ Hz or } f_s = 366\text{ Hz}$
d. $f_s = 425\text{ Hz or } f_s = 377\text{ Hz}$

71. What is the frequency of a stationary sound source if you hear it at 1200.0 Hz while moving towards it at a speed of 50.0 m/s? (Assume speed of sound to be 331 m/s.)
a. 1410 Hz
b. 1380 Hz
c. 1020 Hz
d. 1042 Hz

14.4 Sound Interference and Resonance

72. What is the actual frequency of the wave produced as a result of superposition of two waves?
a. It is the average of the frequencies of the two original waves that were superimposed.
b. It is the difference between the frequencies of the two original waves that were superimposed.
c. It is the product of the frequencies of the two original waves that were superimposed.
d. It is the sum of the frequencies of the two original waves that were superimposed.

73. Can beats be produced through a phenomenon different from resonance? How?
a. No, beats can be produced only by resonance.
b. Yes, beats can be produced by superimposition of any two waves having slightly different frequencies.

74. How is human speech produced?
a. Human speech is produced by shaping the cavity formed by the throat and mouth, the vibration of vocal cords, and using the tongue to adjust the fundamental frequency and combination of overtones.
b. Human speech is produced by shaping the cavity formed by the throat and mouth into a closed pipe and using tongue to adjust the fundamental frequency and combination of overtones.
c. Human speech is produced only by the vibrations of the tongue.
d. Human speech is produced by elongating the vocal cords.

75. What is the possible number of nodes and antinodes along one full wavelength of a standing wave?
a. 2 nodes and 3 antinodes or 2 antinodes and 3 nodes.
b. 2 nodes and 2 antinodes or 3 antinodes and 3 nodes.
c. 3 nodes and 3 antinodes or 2 antinodes and 2 nodes.
d. 6 nodes and 4 antinodes or 6 antinodes and 4 nodes.

76. In a pipe resonator, which frequency will be the least intense of those given below?
a. second overtone frequency
b. first overtone frequency
c. fundamental frequency
d. third overtone frequency

77. A flute is an open-pipe resonator. If a flute is 60 cm long, what is the longest wavelength it can produce?
a. 240 cm
b. 180 cm
c. 60 cm
d. 120 cm

78. What is the frequency of the second overtone of a closed-pipe resonator with a length of 22.0 cm? (Assume the speed of sound is 331 m/s.)
a. 7520 Hz
b. 1510 Hz
c. 376 Hz
d. 1880 Hz

79. An open-pipe resonator has a fundamental frequency of 220 Hz when the speed of sound is 331 m/s. What will its fundamental frequency be when the speed of sound is 350 m/s?
a. 690 Hz
b. 470 Hz
c. 110 Hz
d. 230 Hz

Extended Response

14.1 Speed of Sound, Frequency, and Wavelength

80. How is a human able to hear sounds?
- a. Sound waves cause the eardrum to vibrate. A complicated mechanism converts the vibrations to nerve impulses, which are perceived by the person as sound.
- b. Sound waves cause the ear canal to vibrate. A complicated mechanism converts the vibrations to nerve impulses, which are perceived by the person as sound.
- c. Sound waves transfer electrical impulses to the eardrum. A complicated mechanism converts the electrical impulses to sound.
- d. Sound waves transfer mechanical vibrations to the ear canal, and the eardrum converts them to electrical impulses.

81. Why does sound travel faster in iron than in air even though iron is denser than air?
- a. The density of iron is greater than that of air. However, the rigidity of iron is much greater than that of air. Hence, sound travels faster in it.
- b. The density of iron is greater than that of air. However, the rigidity of iron is much less than that of air. Hence, sound travels faster in it.
- c. The density of iron is greater than that of air. However, the rigidity of iron is equal to that of air. Hence, sound travels faster in it.
- d. The mass of iron is much less than that of air and the rigidity of iron is much greater than that of air. Hence, sound travels faster in it.

82. Is the speed of sound dependent on its frequency?
- a. No
- b. Yes

14.2 Sound Intensity and Sound Level

83. Why is the sound from a tire burst louder than that from a finger snap?
- a. The sound from the tire burst has higher pressure amplitudes, hence it can exert smaller force on the eardrum.
- b. The sound from the tire burst has lower pressure amplitudes, hence it can exert smaller force on the eardrum.
- c. The sound from the tire burst has lower pressure amplitudes, hence it can exert larger force on the ear drum.
- d. The sound from the tire burst has higher pressure amplitudes, hence it can exert larger force on the eardrum.

84. Sound A is 1000 times more intense than Sound B. What will be the difference in decibels in their sound intensity levels?
- a. 5 dB
- b. 10 dB
- c. 3 dB
- d. 30 dB

85. The ratio of the pressure amplitudes of two sound waves traveling through water at 0°C is 4.0. What will be the difference in their sound intensity levels in dB?
- a. 1.2 dB
- b. 6.0 dB
- c. 0.60 dB
- d. 12 dB

86. Which of the following most closely models how sound is produced by the vocal cords?
- a. A person plucks a string.
- b. A person blows over the mouth of a half-filled glass bottle.
- c. A person strikes a hammer against a hard surface.
- d. A person blows through a small slit in a wide, stretched rubber band.

14.3 Doppler Effect and Sonic Booms

87. True or false—The Doppler effect occurs only when the sound source is moving.
- a. False
- b. True

88. True or false—The observed frequency becomes infinite when the source is moving at the speed of sound.
- a. False
- b. True

89. You are driving alongside a train. You hear its horn at a pitch that is lower than the actual frequency. What should you do to match the speed of the train? Why?
- a. In order to match the speed of the train, one would need to increase or decrease the speed of his/her car because a lower pitch means that either the train (the source) is moving away or that you (the observer) are moving away.
- b. In order to match the speed of the train, one would need to drive at a constant speed because a lower pitch means that the train and the car are at the same speed.

14.4 Sound Interference and Resonance

90. How are the beat frequency and the regular frequency of a wave resulting from superposition of two waves different?

a. Beat frequency is the sum of two frequencies and regular frequency is the difference between frequencies of two original waves.
b. Beat frequency is the difference between the constituent frequencies, but the regular frequency is the average of the frequencies of the two original waves.
c. Beat frequency is the sum of two frequencies and regular frequency is the average of frequencies of two original waves.
d. Beat frequency is the average of frequencies of two original waves and regular frequency is the sum of two original frequencies.

91. In the tuning fork and tube experiment, if resonance is formed for $L = \lambda/4$, where L is the length of the tube and λ is the wavelength of the sound wave, can resonance also be formed for a wavelength $\lambda = 4L/9$? Why?
a. The frequency formed is a harmonic and first overtone so resonance will occur.
b. The frequency formed is a harmonic and second overtone so resonance will occur.
c. The frequency formed is a harmonic and third overtone so resonance will occur.
d. The frequency formed is a harmonic and fourth overtone so resonance will occur.

92. True or false—An open-pipe resonator has more overtones than a closed-pipe resonator.
a. False
b. True

93. A flute has finger holes for changing the length of the resonating air column, and therefore, the frequency of the note played. How far apart are two holes that, when closed, play two frequencies that are $300.0\ \text{Hz}$ apart, if the first hole is $20.0\ \text{cm}$ away from the mouthpiece of the flute?
a. $0.31\ \text{m}$
b. $0.24\ \text{m}$
c. $0.04\ \text{m}$
d. $0.11\ \text{m}$

CHAPTER 15
Light

Figure 15.1 Human eyes detect these orange *sea goldie* fish swimming over a coral reef in the blue waters of the Gulf of Eilat, in the Red Sea, using visible light. (credit: David Darom, Wikimedia Commons)

Chapter Outline

15.1 The Electromagnetic Spectrum

15.2 The Behavior of Electromagnetic Radiation

INTRODUCTION The beauty of a coral reef, the warm radiance of sunshine, the sting of sunburn, the X-ray revealing a broken bone, even microwave popcorn—all are brought to us by electromagnetic waves. The list of the various types of electromagnetic waves, ranging from radio transmission waves to nuclear gamma-ray (γ-ray) emissions, is interesting in itself.

Even more intriguing is that all of these different phenomena are manifestations of the same thing—electromagnetic waves (see Figure 15.1). What are electromagnetic waves? How are they created, and how do they travel? How can we understand their widely varying properties? What is the relationship between electric and magnetic effects? These and other questions will be explored.

15.1 The Electromagnetic Spectrum

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Define the electromagnetic spectrum, and describe it in terms of frequencies and wavelengths
- Describe and explain the differences and similarities of each section of the electromagnetic spectrum and the applications of radiation from those sections

Section Key Terms

electric field	electromagnetic radiation (EMR)	magnetic field	Maxwell's equations

The Electromagnetic Spectrum

We generally take light for granted, but it is a truly amazing and mysterious form of energy. Think about it: Light travels to Earth across millions of kilometers of empty space. When it reaches us, it interacts with matter in various ways to generate almost all the energy needed to support life, provide heat, and cause weather patterns. Light is a form of **electromagnetic radiation (EMR)**. The term *light* usually refers to visible light, but this is not the only form of EMR. As we will see, visible light occupies a narrow band in a broad range of types of electromagnetic radiation.

Electromagnetic radiation is generated by a moving electric charge, that is, by an electric current. As you will see when you study electricity, an electric current generates both an **electric field**, **E**, and a **magnetic field**, **B**. These fields are perpendicular to each other. When the moving charge oscillates, as in an alternating current, an EM wave is propagated. Figure 15.2 shows how an electromagnetic wave moves away from the source—indicated by the ~ symbol.

WATCH PHYSICS

Electromagnetic Waves and the Electromagnetic Spectrum

This video, link below, is closely related to the following figure. If you have questions about EM wave properties, the EM spectrum, how waves propagate, or definitions of any of the related terms, the answers can be found in this video (http://www.openstax.org/l/28EMWaves) .

Click to view content (https://www.openstax.org/l/28EMWaves)

GRASP CHECK

In an electromagnetic wave, how are the magnetic field, the electric field, and the direction of propagation oriented to each other?

a. All three are parallel to each other and are along the *x*-axis.
b. All three are mutually perpendicular to each other.
c. The electric field and magnetic fields are parallel to each other and perpendicular to the direction of propagation.
d. The magnetic field and direction of propagation are parallel to each other along the *y*-axis and perpendicular to the electric field.

Virtual Physics

Radio Waves and Electromagnetic Fields

Click to view content (https://www.openstax.org/l/28Radiowaves)

This simulation demonstrates wave propagation. The EM wave is propagated from the broadcast tower on the left, just as in Figure 15.2. You can make the wave yourself or allow the animation to send it. When the wave reaches the antenna on the right, it causes an oscillating current. This is how radio and television signals are transmitted and received.

GRASP CHECK

Where do radio waves fall on the electromagnetic spectrum?

a. Radio waves have the same wavelengths as visible light.
b. Radio waves fall on the high-frequency side of visible light.
c. Radio waves fall on the short-wavelength side of visible light.
d. Radio waves fall on the low-frequency side of visible light.

Figure 15.2 A part of the electromagnetic wave sent out from an oscillating charge at one instant in time. The electric and magnetic fields (**E** and **B**) are in phase, and they are perpendicular to each other and to the direction of propagation. For clarity, the waves are shown only along one direction, but they propagate out in other directions too.

From your study of sound waves, recall these features that apply to all types of waves:

- Wavelength—The distance between two wave crests or two wave troughs, expressed in various metric measures of distance
- Frequency—The number of wave crests that pass a point per second, expressed in hertz (Hz or s^{-1})
- Amplitude: The height of the crest above the null point

As mentioned, electromagnetic radiation takes several forms. These forms are characterized by a range of frequencies. Because frequency is inversely proportional to wavelength, any form of EMR can also be represented by its range of wavelengths. Figure 15.3 shows the frequency and wavelength ranges of various types of EMR. With how many of these types are you familiar?

Figure 15.3 The electromagnetic spectrum, showing the major categories of electromagnetic waves. The range of frequencies and wavelengths is remarkable. The dividing line between some categories is distinct, whereas other categories overlap.

Take a few minutes to study the positions of the various types of radiation on the EM spectrum, above. Sometimes all radiation with frequencies lower than those of visible light are referred to as infrared (IR) radiation. This includes radio waves, which overlap with the frequencies used for media broadcasts of TV and radio signals. The microwave radiation that you see on the diagram is the same radiation that is used in a microwave oven. What we feel as radiant heat is also a form of low-frequency EMR.

All the high-frequency radiation to the right of visible light is sometimes referred to as ultraviolet (UV) radiation. This includes X-rays and gamma (γ) rays. The narrow band that is visible light extends from lower-frequency red light to higher-frequency violet light, thus the terms are *infrared* (below red) and *ultraviolet* (beyond violet).

BOUNDLESS PHYSICS

Maxwell's Equations

The Scottish physicist James Clerk Maxwell (1831–1879) is regarded widely to have been the greatest theoretical physicist of the

nineteenth century. Although he died young, Maxwell not only formulated a complete electromagnetic theory, represented by **Maxwell's equations**, he also developed the kinetic theory of gases, and made significant contributions to the understanding of color vision and the nature of Saturn's rings.

Maxwell brought together all the work that had been done by brilliant physicists, such as Ørsted, Coulomb, Ampere, Gauss, and Faraday, and added his own insights to develop the overarching theory of electromagnetism. Maxwell's equations are paraphrased here in words because their mathematical content is beyond the level of this text. However, the equations illustrate how apparently simple mathematical statements can elegantly unite and express a multitude of concepts—why mathematics is the language of science.

Maxwell's Equations

1. Electric field lines originate on positive charges and terminate on negative charges. The electric field is defined as the force per unit charge on a test charge, and the strength of the force is related to the electric constant, ε_0.
2. Magnetic field lines are continuous, having no beginning or end. No magnetic monopoles are known to exist. The strength of the magnetic force is related to the magnetic constant, μ_0.
3. A changing magnetic field induces an electromotive force (emf) and, hence, an electric field. The direction of the emf opposes the change, changing direction of the magnetic field.
4. Magnetic fields are generated by moving charges or by changing electric fields.

Maxwell's complete theory shows that electric and magnetic forces are not separate, but different manifestations of the same thing—the electromagnetic force. This classical unification of forces is one motivation for current attempts to unify the four basic forces in nature—the gravitational, electromagnetic, strong nuclear, and weak nuclear forces. The weak nuclear and electromagnetic forces have been unified, and further unification with the strong nuclear force is expected; but, the unification of the gravitational force with the other three has proven to be a real head-scratcher.

One final accomplishment of Maxwell was his development in 1855 of a process that could produce color photographic images. In 1861, he and photographer Thomas Sutton worked together on this process. The color image was achieved by projecting red, blue, and green light through black-and-white photographs of a tartan ribbon, each photo itself exposed in different-colored light. The final image was projected onto a screen (see Figure 15.4).

Figure 15.4 Maxwell and Sutton's photograph of a colored ribbon. This was the first durable color photograph. The plaid tartan of the Scots made a *colorful* photographic subject.

GRASP CHECK

Describe electromagnetic force as explained by Maxwell's equations.

a. According to Maxwell's equations, electromagnetic force gives rise to electric force and magnetic force.
b. According to Maxwell's equations, electric force and magnetic force are different manifestations of electromagnetic force.
c. According to Maxwell's equations, electric force is the cause of electromagnetic force.
d. According to Maxwell's equations, magnetic force is the cause of electromagnetic force.

Characteristics of Electromagnetic Radiation

All the EM waves mentioned above are basically the same form of radiation. They can all travel across empty space, and they all

travel at the speed of light in a vacuum. The basic difference between types of radiation is their differing frequencies. Each frequency has an associated wavelength. As frequency increases across the spectrum, wavelength decreases. Energy also increases with frequency. Because of this, higher frequencies penetrate matter more readily. Some of the properties and uses of the various EM spectrum bands are listed in Table 15.1.

Types of EM Waves	Production	Applications	Life Sciences Aspect	Issues
Radio and TV	Accelerating charges	Communications, remote controls	MRI	Requires controls for band use
Microwaves	Accelerating charges & thermal agitation	Communications, microwave ovens, radar	Deep heating	Cell phone use
Infrared	Thermal agitation & electronic transitions	Thermal imaging, heating	Absorption by atmosphere	Greenhouse effect
Visible Light	Thermal agitation & electronic transitions	All pervasive	Photosynthesis, human vision	
Ultraviolet	Thermal agitation & electronic transitions	Sterilization, slowing abnormal growth of cells	Vitamin D production	Ozone depletion, causes cell damage
X-rays	Inner electronic transitions & fast collisions	Medical, security	Medical diagnosis, cancer therapy	Causes cell damage
Gamma Rays	Nuclear decay	Nuclear medicine, security	Medical diagnosis, cancer therapy	Causes cell damage, radiation damage

Table 15.1 Electromagnetic Waves This table shows how each type of EM radiation is produced, how it is applied, as well as environmental and health issues associated with it.

The narrow band of visible light is a combination of the colors of the rainbow. Figure 15.5 shows the section of the EM spectrum that includes visible light. The frequencies corresponding to these wavelengths are $4.0 \times 10^{14}\ \mathrm{s}^{-1}$ at the red end to $7.9 \times 10^{14}\ \mathrm{s}^{-1}$ at the violet end. This is a very narrow range, considering that the EM spectrum spans about 20 orders of magnitude.

Figure 15.5 A small part of the electromagnetic spectrum that includes its visible components. The divisions between infrared, visible, and ultraviolet are not perfectly distinct, nor are the divisions between the seven rainbow colors

TIPS FOR SUCCESS

Wavelengths of visible light are often given in nanometers, nm. One nm equals 10^{-9} m. For example, yellow light has a wavelength of about 600 nm, or 6×10^{-7} m.

As a child, you probably learned the color wheel, shown on the left in Figure 15.6. It helps if you know what color results when you mix different colors of paint together. Mixing two of the primary *pigment* colors—magenta, yellow, or cyan—together results in a secondary color. For example, mixing cyan and yellow makes green. This is called *subtractive* color mixing. Mixing different colors of *light* together is quite different. The diagram on the right shows *additive* color mixing. In this case, the primary colors are red, green, and blue, and the secondary colors are cyan, magenta, and yellow. Mixing pigments and mixing light are different because materials absorb light by a different set of rules than does the perception of light by the eye. Notice that, when all colors are subtracted, the result is no color, or black. When all colors are added, the result is white light. We see the reverse of this when white sunlight is separated into the visible spectrum by a prism or by raindrops when a rainbow appears in the sky.

Figure 15.6 Mixing colored pigments follows the subtractive color wheel, and mixing colored light follows the additive color wheel.

Virtual Physics

Color Vision

Click to view content (https://www.openstax.org/l/28Colorvision)

This video demonstrates additive color and color filters. Try all the settings except *Photons*.

GRASP CHECK

Explain why only light from a blue bulb passes through the blue filter.

a. A blue filter absorbs blue light.
b. A blue filter reflects blue light.
c. A blue filter absorbs all visible light other than blue light.
d. A blue filter reflects all of the other colors of light and absorbs blue light.

LINKS TO PHYSICS

Animal Color Perception

The physics of color perception has interesting links to zoology. Other animals have very different views of the world than humans, especially with respect to which colors can be seen. Color is detected by cells in the eye called *cones*. Humans have three cones that are sensitive to three different ranges of electromagnetic wavelengths. They are called red, blue, and green cones, although these colors do not correspond exactly to the centers of the three ranges. The ranges of wavelengths that each cone detects are red, 500 to 700 nm; green, 450 to 630 nm; and blue, 400 to 500 nm.

Most primates also have three kinds of cones and see the world much as we do. Most mammals other than primates only have two cones and have a less colorful view of things. Dogs, for example see blue and yellow, but are color blind to red and green. You might think that *simpler* species, such as fish and insects, would have less sophisticated vision, but this is not the case. Many birds, reptiles, amphibians, and insects have four or five different cones in their eyes. These species don't have a wider range of perceived colors, but they see more hues, or combinations of colors. Also, some animals, such as bees or rattlesnakes, see a

range of colors that is as broad as ours, but shifted into the ultraviolet or infrared.

These differences in color perception are generally adaptations that help the animals survive. Colorful tropical birds and fish display some colors that are too subtle for us to see. These colors are believed to play a role in the species mating rituals. Figure 15.7 shows the colors visible and the color range of vision in humans, bees, and dogs.

Figure 15.7 Humans, bees, and dogs see colors differently. Dogs see fewer colors than humans, and bees see a different range of colors.

GRASP CHECK

The belief that bulls are enraged by seeing the color red is a misconception. What did you read in this Links to Physics that shows why this belief is incorrect?

a. Bulls are color-blind to every color in the spectrum of colors.
b. Bulls are color-blind to the blue colors in the spectrum of colors.
c. Bulls are color-blind to the red colors in the spectrum of colors.
d. Bulls are color-blind to the green colors in the spectrum of colors.

Humans have found uses for every part of the electromagnetic spectrum. We will take a look at the uses of each range of frequencies, beginning with visible light. Most of our uses of visible light are obvious; without it our interaction with our surroundings would be much different. We might forget that nearly all of our food depends on the photosynthesis process in plants, and that the energy for this process comes from the visible part of the spectrum. Without photosynthesis, we would also have almost no oxygen in the atmosphere.

The low-frequency, infrared region of the spectrum has many applications in media broadcasting. Television, radio, cell phone, and remote-control devices all broadcast and/or receive signals with these wavelengths. AM and FM radio signals are both low-frequency radiation. They are in different regions of the spectrum, but that is not their basic difference. AM and FM are abbreviations for *amplitude modulation* and *frequency modulation*. Information in AM signals has the form of changes in *amplitude* of the radio waves; information in FM signals has the form of changes in wave *frequency*.

Another application of long-wavelength radiation is found in microwave ovens. These appliances cook or warm food by irradiating it with EM radiation in the microwave frequency range. Most kitchen microwaves use a frequency of 2.45×10^9 Hz. These waves have the right amount of energy to cause polar molecules, such as water, to rotate faster. Polar molecules are those that have a partial charge separation. The rotational energy of these molecules is given up to surrounding matter as heat. The first microwave ovens were called *Radaranges* because they were based on radar technology developed during World War II.

Radar uses radiation with wavelengths similar to those of microwaves to detect the location and speed of distant objects, such as airplanes, weather formations, and motor vehicles. Radar information is obtained by receiving and analyzing the echoes of microwaves reflected by an object. The speed of the object can be measured using the Doppler shift of the returning waves. This is the same effect you learned about when you studied sound waves. Like sound waves, EM waves are shifted to higher frequencies by an object moving toward an observer, and to lower frequencies by an object moving away from the observer. Astronomers use this same Doppler effect to measure the speed at which distant galaxies are moving away from us. In this case, the shift in frequency is called the *red shift*, because visible frequencies are shifted toward the lower-frequency, red end of the spectrum.

Exposure to any radiation with frequencies greater than those of visible light carries some health hazards. All types of radiation in this range are known to cause cell damage. The danger is related to the high energy and penetrating ability of these EM waves. The likelihood of being harmed by any of this radiation depends largely on the amount of exposure. Most people try to reduce exposure to UV radiation from sunlight by using sunscreen and protective clothing. Physicians still use X-rays to diagnose medical problems, but the intensity of the radiation used is extremely low. Figure 15.8 shows an X-ray image of a patient's chest cavity.

One medical-imaging technique that involves no danger of exposure is magnetic resonance imaging (MRI). MRI is an important imaging and research tool in medicine, producing highly detailed two- and three-dimensional images. Radio waves are broadcast, absorbed, and reemitted in a resonance process that is sensitive to the density of nuclei, usually hydrogen nuclei—protons.

Figure 15.8 This shadow X-ray image shows many interesting features, such as artificial heart valves, a pacemaker, and wires used to close the sternum. (credit: P.P. Urone)

Check Your Understanding

1. Identify the fields produced by a moving charged particle.
   a. Both an electric field and a magnetic field will be produced.
   b. Neither a magnetic field nor an electric field will be produced.
   c. A magnetic field, but no electric field will be produced.
   d. Only the electric field, but no magnetic field will be produced.

2. X-rays carry more energy than visible light. Compare the frequencies and wavelengths of these two types of EM radiation.
   a. Visible light has higher frequencies and shorter wavelengths than X-rays.
   b. Visible light has lower frequencies and shorter wavelengths than X-rays.
   c. Visible light has higher frequencies and longer wavelengths than X-rays.
   d. Visible light has lower frequencies and longer wavelengths than X-rays.

3. How does wavelength change as frequency increases across the EM spectrum?
   a. The wavelength increases.
   b. The wavelength first increases and then decreases.
   c. The wavelength first decreases and then increases.
   d. The wavelength decreases.

4. Why are X-rays used in imaging of broken bones, rather than radio waves?
   a. X-rays have higher penetrating energy than radio waves.
   b. X-rays have lower penetrating energy than radio waves.
   c. X-rays have a lower frequency range than radio waves.
   d. X-rays have longer wavelengths than radio waves.

5. Identify the fields that make up an electromagnetic wave.
   a. both an electric field and a magnetic field
   b. neither a magnetic field nor an electric field

c. only a magnetic field, but no electric field
d. only an electric field, but no magnetic field

15.2 The Behavior of Electromagnetic Radiation

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Describe the behavior of electromagnetic radiation
- Solve quantitative problems involving the behavior of electromagnetic radiation

Section Key Terms

illuminance	interference	lumens	luminous flux	lux	polarized light

Types of Electromagnetic Wave Behavior

In a vacuum, all electromagnetic radiation travels at the same incredible speed of 3.00×10^8 m/s, which is equal to 671 million miles per hour. This is one of the fundamental physical constants. It is referred to as the speed of light and is given the symbol c. The space between celestial bodies is a near vacuum, so the light we see from the Sun, stars, and other planets has traveled here at the speed of light. Keep in mind that all EM radiation travels at this speed. All the different wavelengths of radiation that leave the Sun make the trip to Earth in the same amount of time. That trip takes 8.3 minutes. Light from the nearest star, besides the Sun, takes 4.2 years to reach Earth, and light from the nearest galaxy—a dwarf galaxy that orbits the Milky Way—travels 25,000 years on its way to Earth. You can see why we call very long distances *astronomical*.

When light travels through a physical medium, its speed is always less than the speed of light. For example, light travels in water at three-fourths the value of c. In air, light has a speed that is just slightly slower than in empty space: 99.97 percent of c. Diamond slows light down to just 41 percent of c. When light changes speeds at a boundary between media, it also changes direction. The greater the difference in speeds, the more the path of light bends. In other chapters, we look at this bending, called *refraction*, in greater detail. We introduce refraction here to help explain a phenomenon called thin-film interference.

Have you ever wondered about the rainbow colors you often see on soap bubbles, oil slicks, and compact discs? This occurs when light is both refracted by and reflected from a very thin film. The diagram shows the path of light through such a thin film. The symbols n_1, n_2, and n_3 indicate that light travels at different speeds in each of the three materials. Learn more about this topic in the chapter on diffraction and interference.

Figure 15.9 shows the result of thin film interference on the surface of soap bubbles. Because ray 2 travels a greater distance, the two rays become *out of phase*. That is, the crests of the two emerging waves are no longer moving together. This causes **interference**, which reinforces the intensity of the wavelengths of light that create the bands of color. The color bands are separated because each color has a different wavelength. Also, the thickness of the film is not uniform, and different thicknesses cause colors of different wavelengths to interfere in different places. Note that the film must be very, very thin—somewhere in the vicinity of the wavelengths of visible light.

Figure 15.9 Light striking a thin film is partially reflected (ray 1) and partially refracted at the top surface. The refracted ray is partially reflected at the bottom surface and emerges as ray 2. These rays will interfere in a way that depends on the thickness of the film and the indices of refraction of the various media.

You have probably experienced how polarized sunglasses reduce glare from the surface of water or snow. The effect is caused by the wave nature of light. Looking back at , we see that the electric field moves in only one direction perpendicular to the direction of propagation. Light from most sources vibrates in all directions perpendicular to propagation. Light with an electric field that vibrates in only one direction is called *polarized*. A diagram of **polarized light** would look like .

Polarized glasses are an example of a polarizing filter. These glasses absorb most of the horizontal light waves and transmit the vertical waves. This cuts down glare, which is caused by horizontal waves. Figure 15.10 shows how waves traveling along a rope can be used as a model of how a polarizing filter works. The oscillations in one rope are in a vertical plane and are said to be vertically polarized. Those in the other rope are in a horizontal plane and are horizontally polarized. If a vertical slit is placed on the first rope, the waves pass through. However, a vertical slit blocks the horizontally polarized waves. For EM waves, the direction of the electric field oscillation is analogous to the disturbances on the ropes.

Figure 15.10 The transverse oscillations in one rope are in a vertical plane, and those in the other rope are in a horizontal plane. The first is said to be vertically polarized, and the other is said to be horizontally polarized. Vertical slits pass vertically polarized waves and block horizontally polarized waves.

Light can also be polarized by reflection. Most of the light reflected from water, glass, or any highly reflective surface is polarized horizontally. Figure 15.11 shows the effect of a polarizing lens on light reflected from the surface of water.

(a) (b)

Figure 15.11 These two photographs of a river show the effect of a polarizing filter in reducing glare in light reflected from the surface of water. Part (b) of this figure was taken with a polarizing filter and part (a) was taken without. As a result, the reflection of clouds and sky observed in part (a) is not observed in part (b). Polarizing sunglasses are particularly useful on snow and water.

WATCH PHYSICS

Polarization of Light, Linear and Circular

This video explains the polarization of light in great detail. Before viewing the video, look back at the drawing of an electromagnetic wave from the previous section. Try to visualize the two-dimensional drawing in three dimensions.

Click to view content (https://www.openstax.org/l/28Polarization)

GRASP CHECK

How do polarized glasses reduce glare reflected from the ocean?

a. They block horizontally polarized and vertically polarized light.
b. They are transparent to horizontally polarized and vertically polarized light.
c. They block horizontally polarized rays and are transparent to vertically polarized rays.
d. They are transparent to horizontally polarized light and block vertically polarized light.

Snap Lab

Polarized Glasses

- EYE SAFETY—Looking at the Sun directly can cause permanent eye damage. Avoid looking directly at the Sun.
- two pairs of polarized sunglasses
  OR
- two lenses from one pair of polarized sunglasses

Procedure

1. Look through both or either polarized lens and record your observations.
2. Hold the lenses, one in front of the other. Hold one lens stationary while you slowly rotate the other lens. Record your observations, including the relative angles of the lenses when you make each observation.
3. Find a reflective surface on which the Sun is shining. It could be water, glass, a mirror, or any other similar smooth surface. The results will be more dramatic if the sunlight strikes the surface at a sharp angle.
4. Observe the appearance of the surface with your naked eye and through one of the polarized lenses.
5. Observe any changes as you slowly rotate the lens, and note the angles at which you see changes.

GRASP CHECK

If you buy sunglasses in a store, how can you be sure that they are polarized?

a. When one pair of sunglasses is placed in front of another and rotated in the plane of the body, the light passing through the sunglasses will be blocked at two positions due to refraction of light.
b. When one pair of sunglasses is placed in front of another and rotated in the plane of the body, the light passing through the sunglasses will be blocked at two positions due to reflection of light.
c. When one pair of sunglasses is placed in front of another and rotated in the plane of the body, the light passing through the sunglasses will be blocked at two positions due to the polarization of light.
d. When one pair of sunglasses is placed in front of another and rotated in the plane of the body, the light passing through the sunglasses will be blocked at two positions due to the bending of light waves.

Quantitative Treatment of Electromagnetic Waves

We can use the speed of light, *c*, to carry out several simple but interesting calculations. If we know the distance to a celestial object, we can calculate how long it takes its light to reach us. Of course, we can also make the reverse calculation if we know the time it takes for the light to travel to us. For an object at a very great distance from Earth, it takes many years for its light to reach us. This means that we are looking at the object as it existed in the distant past. The object may, in fact, no longer exist. Very large distances in the universe are measured in light years. One light year is the distance that light travels in one year, which is 9.46×10^{12} kilometers or 5.88×10^{12} miles (...and 10^{12} is a trillion!).

A useful equation involving *c* is

$$c = f\lambda \tag{15.1}$$

where *f* is frequency in Hz, and λ is wavelength in meters.

WORKED EXAMPLE

Frequency and Wavelength Calculation

For example, you can calculate the frequency of yellow light with a wavelength of 6.00×10^{-7} m.

STRATEGY

Rearrange the equation to solve for frequency.

$$f = \frac{c}{\lambda} \tag{15.2}$$

Solution

Substitute the values for the speed of light and wavelength into the equation.

$$f = \frac{3.00 \times 10^{8}\ \text{m/s}}{6.00 \times 10^{-7}\ \text{m}} = 5.00 \times 10^{14}\ \text{s}^{-1} = 5.00 \times 10^{14}\ \text{Hz} \tag{15.3}$$

Discussion

Manipulating exponents of 10 in a fraction can be tricky. Be sure you keep track of the + and – exponents correctly. Checking back to the diagram of the electromagnetic spectrum in the previous section shows that 10^{14} is a reasonable order of magnitude for the frequency of yellow light.

The frequency of a wave is proportional to the energy the wave carries. The actual proportionality constant will be discussed in a later chapter. Since frequency is inversely proportional to wavelength, we also know that wavelength is inversely proportional to energy. Keep these relationships in mind as general rules.

The rate at which light is radiated from a source is called **luminous flux**, *P*, and it is measured in **lumens** (lm). Energy-saving light bulbs, which provide more luminous flux for a given use of electricity, are now available. One of these bulbs is called a *compact fluorescent lamp*; another is an *LED* (light-emitting diode) bulb. If you wanted to replace an old incandescent bulb with an energy saving bulb, you would want the new bulb to have the same brightness as the old one. To compare bulbs accurately, you would need to compare the lumens each one puts out. Comparing wattage—that is, the electric power used—would be

misleading. Both wattage and lumens are stated on the packaging.

The luminous flux of a bulb might be 2,000 lm. That accounts for all the light radiated in all directions. However, what we really need to know is how much light falls on an object, such as a book, at a specific distance. The number of lumens per square meter is called **illuminance**, and is given in units of **lux** (lx). Picture a light bulb in the middle of a sphere with a 1-m radius. The total surface of the sphere equals $4\pi r^2$ m^2. The illuminance then is given by

$$\text{illuminance} = \frac{P}{4\pi r^2}. \tag{15.4}$$

What happens if the radius of the sphere is increased 2 m? The illuminance is now only one-fourth as great, because the r^2 term in the denominator is 4 instead of 1. Figure 15.12 shows how illuminance decreases with the inverse square of the distance.

Figure 15.12 The diagram shows why the illuminance varies inversely with the square of the distance from a source of light.

WORKED EXAMPLE

Calculating Illuminance

A woman puts a new bulb in a floor lamp beside an easy chair. If the luminous flux of the bulb is rated at 2,000 lm, what is the illuminance on a book held 2.00 m from the bulb?

STRATEGY

Choose the equation and list the knowns.

Equation: $\text{illuminance} = \frac{P}{4\pi r^2}$

$P = 2{,}000$ lm

$\pi = 3.14$

$r = 2.00$ m

Solution

Substitute the known values into the equation.

$$\begin{aligned}\text{illuminance} &= \frac{P}{4\pi r^2}\\ &= \frac{2{,}000\ \text{lm}}{4(3.14)(2.00)^2\ \text{m}^2}\\ &= 39.8\ \text{lm/m}^2\\ &= 39.8\ \text{lx}\end{aligned}$$

Discussion

Try some other distances to illustrate how greatly light fades with distance from its source. For example, at 3 m the illuminance is only 17.7 lux. Parents often scold children for reading in light that is too dim. Instead of shouting, "You'll ruin your eyes!" it might be better to explain the inverse square law of illuminance to the child.

Practice Problems

6. Red light has a wavelength of 7.0 × 10–7 m and a frequency of 4.3 × 1014 Hz. Use these values to calculate the speed of light in a vacuum.
   a. 3 × 1020 m/s
   b. 3 × 1015 m/s
   c. 3 × 1014 m/s
   d. 3 × 108 m/s

7. A light bulb has a luminous flux of 942 lumens. What is the illuminance on a surface $3.00\ \text{m}$ from the bulb when it is lit?
   a. $33.32\ \text{lx}$
   b. $26.15\ \text{lx}$
   c. $2.77\ \text{lx}$
   d. $8.33\ \text{lx}$

Check Your Understanding

8. Give an example of a place where light travels at the speed of 3.00×10^8 m/s.
   a. outer space
   b. water
   c. Earth's atmosphere
   d. quartz glass

9. Explain in terms of distances and the speed of light why it is currently very unlikely that humans will visit planets that circle stars other than our Sun.
   a. The spacecrafts used for travel are very heavy and thus very slow.
   b. Spacecrafts do not have a constant source of energy to run them.
   c. If a spacecraft could attain a maximum speed equal to that of light, it would still be too slow to cover astronomical distances.
   d. Spacecrafts can attain a maximum speed equal to that of light, but it is difficult to locate planets around stars.

KEY TERMS

electric field a field that tells us the force per unit charge at all locations in space around a charge distribution

electromagnetic radiation (EMR) radiant energy that consists of oscillating electric and magnetic fields

illuminance number of lumens per square meter, given in units of lux (lx)

interference increased or decreased light intensity caused by the phase differences between waves

lumens unit of measure for luminous flux

luminous flux rate at which light is radiated from a source

lux unit of measure for illuminance

magnetic field the directional lines around a magnetic material that indicates the direction and magnitude of the magnetic force

Maxwell's equations equations that describe the interrelationship between electric and magnetic fields, and how these fields combine to form electromagnetic radiation

polarized light light whose electric field component vibrates in a specific plane

SECTION SUMMARY

15.1 The Electromagnetic Spectrum

- The electromagnetic spectrum is made up of a broad range of frequencies of electromagnetic radiation.
- All frequencies of EM radiation travel at the same speed in a vacuum and consist of an electric field and a magnetic field. The types of EM radiation have different frequencies and wavelengths, and different energies and penetrating ability.

15.2 The Behavior of Electromagnetic Radiation

- EM radiation travels at different speeds in different media, produces colors on thin films, and can be polarized to oscillate in only one direction.
- Calculations can be based on the relationship among the speed, frequency, and wavelength of light, and on the relationship among luminous flux, illuminance, and distance.

KEY EQUATIONS

15.2 The Behavior of Electromagnetic Radiation

frequency and wavelength $c = f\lambda$

illuminance $\text{illuminance} = \frac{P}{4\pi r^2}$

CHAPTER REVIEW

Concept Items

15.1 The Electromagnetic Spectrum

1. Use the concepts on which Maxwell's equations are based to explain why a compass needle is deflected when the compass is brought near a wire that is carrying an electric current.
   a. The charges in the compass needle and the charges in the electric current have interacting electric fields, causing the needle to deflect.
   b. The electric field from the moving charges in the current interacts with the magnetic field of the compass needle, causing the needle to deflect.
   c. The magnetic field from the moving charges in the current interacts with the electric field of the compass needle, causing the needle to deflect.
   d. The moving charges in the current produce a magnetic field that interacts with the compass needle's magnetic field, causing the needle to deflect.

2. Consider these colors of light: yellow, blue, and red. Part A. Put these light waves in order according to wavelength, from shortest wavelength to longest wavelength. Part B. Put these light waves in order according to frequency, from lowest frequency to highest frequency.
   a. wavelength: blue, yellow, red
      frequency: blue, yellow, red
   b. wavelength: red, yellow, blue
      frequency: red, yellow, blue
   c. wavelength: red, yellow, blue
      frequency: blue, yellow, red
   d. wavelength: blue, yellow, red
      frequency: red, yellow, blue

3. Describe the location of gamma rays on the electromagnetic spectrum.

a. At the high-frequency and long-wavelength end of the spectrum
b. At the high-frequency and short-wavelength end of the spectrum
c. At the low-frequency and long-wavelength end of the spectrum
d. At the low-frequency and short-wavelength end of the spectrum

4. In which region of the electromagnetic spectrum would you find radiation that is invisible to the human eye and has low energy?
   a. Long-wavelength and high-frequency region
   b. Long-wavelength and low-frequency region
   c. Short-wavelength and high-frequency region
   d. Short-wavelength and low-frequency region

15.2 The Behavior of Electromagnetic Radiation

5. Light travels at different speeds in different media. Put these media in order, from the slowest light speed to the fastest light speed: air, diamond, vacuum, water.
   a. diamond, water, air, vacuum
   b. vacuum, diamond, air, water
   c. diamond, air, water, vacuum
   d. air, diamond, water, vacuum

6. Visible light has wavelengths in the range of about 400 to 800 nm. What does this indicate about the approximate thickness of the wall of a soap bubble? Explain your answer.
   a. The thickness of the bubble wall is ten times that of the wavelength of light.
   b. The thickness of the bubble wall is similar to that of the wavelength of light.
   c. The thickness of the bubble wall is half the wavelength of light.
   d. The thickness of the bubble wall equals the cube of the wavelength of light.

7. Bright sunlight is reflected from an icy pond. You look at the glare of the reflected light through polarized glasses. When you take the glasses off, rotate them 90°, and look through one of the lenses again, the light you see becomes brighter. Explain why the light you see changes.
   a. The glass blocks horizontally polarized light, and the light reflected from the icy pond is, in part, polarized horizontally.
   b. The glass blocks vertically polarized light, and the light reflected from the icy pond is, in part, polarized vertically.
   c. The glass allows horizontally polarized light to pass, and the light reflected from the icy pond is, in part, polarized vertically.
   d. The glass allows horizontally polarized light to pass, and the light reflected from the icy pond is, in part, polarized horizontally.

Critical Thinking Items

15.1 The Electromagnetic Spectrum

8. Standing in front of a fire, we can sense both its heat and its light. How are the light and heat radiated by the fire the same, and how are they different?
   a. Both travel as waves, but only light waves are a form of electromagnetic radiation.
   b. Heat and light are both forms of electromagnetic radiation, but light waves have higher frequencies.
   c. Heat and light are both forms of electromagnetic radiation, but heat waves have higher frequencies.
   d. Heat and light are both forms of electromagnetic radiation, but light waves have higher wavelengths.

9. Light shines on a picture of the subtractive color wheel. The light is a mixture of red, blue, and green light.
   Part A—Which part of the color wheel will look blue? Explain in terms of absorbed and reflected light.
   Part B—Which part of the color wheel will look yellow? Explain in terms of absorbed and reflected light.
   a. A. The yellow section of the wheel will look blue because it will reflect blue light and absorb red and green.
      B. The blue section of the wheel will look yellow because it will reflect red and green light and absorb blue.
   b. A. The blue section of the wheel will look blue because it will absorb blue light and reflect red and green.
      B. The yellow section of the wheel will look yellow because it will absorb red and green light and reflect blue.
   c. A. The yellow section of the wheel will look blue because it will absorb blue light and reflect red and green.
      B. The blue section of the wheel will look yellow because it will absorb red and green light and reflect blue.
   d. A. The blue section of the wheel will look blue because it will reflect blue light and absorb red and green.
      B. The yellow section of the wheel will look yellow because it will reflect red and green light and absorb blue.

10. Part A. When you stand in front of an open fire, you can sense light waves with your eyes. You sense another type of electromagnetic radiation as heat. What is this other type of radiation?
Part B. How is this other type of radiation different front light waves?

a. A. X-rays
   B. The X-rays have higher frequencies and shorter wavelengths than the light waves.

b. A. X-rays
   B. The X-rays have lower frequencies and longer wavelengths than the light waves.

c. A. infrared rays
   B. The infrared rays have higher frequencies and shorter wavelengths than the light waves.

d. A. infrared rays
   B. The infrared rays have lower frequencies and longer wavelengths than the light waves.

11. Overexposure to this range of EM radiation is dangerous, and yet it is used by doctors to diagnose medical problems.
Part A—Identify the type of radiation.
Part B—Locate the position of this radiation on the EM spectrum by comparing its frequency and wavelength to visible light.
Part C—Explain why this radiation is both dangerous and therapeutic in terms of its energy, based on your answer to Part B.

a. A. X-rays
   B. X-rays have shorter wavelengths ($1 \times 10^{-8} - 5 \times 10^{-12}$ m) and higher frequencies ($3 \times 10^{16} - 6 \times 10^{19}$ Hz) than visible light ($7.5 \times 10^{-7} - 4.0 \times 10^{-7}$ m; $4.0 \times 10^{14} - 7.5 \times 10^{14}$ Hz).
   C. X-rays have low energies because of their high frequencies, and so can penetrate matter to greater depths.
b. A. X-rays
   B. X-rays have shorter wavelengths ($1 \times 10^{-8} - 5 \times 10^{-12}$ m) and higher frequencies ($3 \times 10^{10} - 6 \times 10^{13}$ Hz) than visible light ($7.5 \times 10^{-7} - 4.0 \times 10^{-7}$ m; $4.0 \times 10^{14} - 7.5 \times 10^{14}$ Hz).
   C. X-rays have low energies because of their low frequencies, and so can penetrate matter to greater depths.
c. A. X-rays B. X-rays have longer wavelengths ($1 \times 10^{-6} - 5 \times 10^{-7}$ m) and higher frequencies ($3 \times 10^{15} - 6 \times 10^{15}$ Hz) than visible light ($7.5 \times 10^{-7} - 4.0 \times 10^{-7}$ m; $4.0 \times 10^{14} - 7.5 \times 10^{14}$ Hz).
   C. X-rays have high energies because of their high frequencies, and therefore can penetrate matter to greater depths.
d. A. X-rays
   B. X-rays have shorter wavelengths ($1 \times 10^{-8} - 5 \times 10^{-12}$ m) and higher frequencies ($3 \times 10^{16} - 6 \times 10^{19}$ Hz) than visible light ($7.5 \times 10^{-7} - 4.0 \times 10^{-7}$ m; $4.0 \times 10^{14} - 7.5 \times 10^{14}$ Hz).
   C. X-rays have high energies because of their high frequencies, and so can penetrate matter to greater depths.

15.2 The Behavior of Electromagnetic Radiation

12. Explain how thin-film interference occurs. Discuss in terms of the meaning of interference and the pathways of light waves.

a. For a particular thickness of film, light of a given wavelength that reflects from the outer and inner film surfaces is completely in phase, and so undergoes constructive interference.
b. For a particular thickness of film, light of a given wavelength that reflects from the outer and inner surfaces is completely in phase, and so undergoes destructive interference.
c. For a particular thickness of film, light of a given wavelength that reflects from the outer and inner film surfaces is completely out of phase, and so undergoes constructive interference.
d. For a particular thickness of film, light of a given wavelength that reflects from the outer and inner film surfaces is completely out of phase, and so undergoes no interference.

13. When you move a rope up and down, waves are created. If the waves pass through a slot, they will be affected differently, depending on the orientation of the slot. Using the rope waves and the slot as a model, explain how polarizing glasses affect light waves.

a. If the wave—electric field—is vertical and slit—polarizing molecules in the glass—is horizontal, the wave will pass.
b. If the wave—electric field— is vertical and slit—polarizing molecules in the glass—is vertical, the wave will not pass.
c. If the wave—electric field—is horizontal and slit—polarizing molecules in the glass—is horizontal, the wave will pass.
d. If the wave—electric field—is horizontal and slit—polarizing molecules in the glass—is horizontal, the wave will not pass.

Problems

15.2 The Behavior of Electromagnetic Radiation

14. Visible light has a range of wavelengths from about 400 nm to 800 nm . What is the range of frequencies for visible light?

a. 3.75×10^{6} Hz to 7.50×10^{6} Hz
b. 3.75 Hz to 7.50 Hz
c. 3.75×10^{-7} Hz to 7.50×10^{-7} Hz
d. 3.75×10^{14} Hz to 7.50×10^{14} Hz

15. Light travels through the wall of a soap bubble that is 600 nm thick and is reflected from the inner surface back into the air. Assume the bubble wall is mostly water and that light travels in water at 75 percent of the speed of light in vacuum. How many seconds behind will the light reflected from the inner surface arrive compared to the light that was reflected from the outer surface?

a. 4.0×10^{-8} s
b. 5.3×10^{-6} s
c. 2.65×10^{-15} s
d. 5.3×10^{-15} s

Performance Task

15.2 The Behavior of Electromagnetic Radiation

16. Design an experiment to observe the phenomenon of thin-film interference. Observe colors of visible light, and relate each color to its corresponding wavelength. Comparison with the magnitudes of visible light wavelength will give an appreciation of just how very thin a thin film is. Thin-film interference has a number of practical applications, such as anti-reflection coatings and optical filters. Thin films used in filters can be designed to reflect or transmit specific wavelengths of light. This is done by depositing a film one molecular layer at a time from a vapor, thus allowing the thickness of the film to be exactly controlled.

- EYE SAFETY—Chemicals in this lab are poisonous if ingested. If chemicals are ingested, inform your teacher immediately.
- FUMES—Certain chemicals or chemical reactions in this lab create a vapor that is harmful if inhaled. Follow your teacher's instructions for the use of fume hoods and other safety apparatus designed to prevent fume inhalation. Never smell or otherwise breath in any chemicals or vapors in the lab.
- FLAMMABLE—Chemicals in this lab are highly flammable and can ignite, especially if exposed to a spark or open flame. Follow your teacher's instructions carefully on how to handle flammable chemicals. Do not expose any chemical to a flame or other heat source unless specifically instructed by your teacher.
- HAND WASHING—Some materials may be hazardous if in extended contact with the skin. Be sure to wash your hands with soap after handling and disposing of these materials during the lab.
- WASTE—Some things in this lab are hazardous and need to be disposed of properly. Follow your teacher's instructions for disposal of all items.

- A large flat tray with raised sides, such as a baking tray
- Small volumes of motor oil, lighter fluid or a penetrating oil of the type used to loosen rusty bolts, and cooking oil
- Water
- A camera

a. Thin-film interference causes colors to appear on the surface of a thin transparent layer. Do you expect to see a pattern to the colors?
b. How could you make a permanent record of your observations?
c. What data would you need to look up to help explain any patterns that you see?
d. What could explain colors failing to appear under some conditions?

TEST PREP

Multiple Choice

15.1 The Electromagnetic Spectrum

17. Which type of EM radiation has the shortest wavelengths?

a. gamma rays
b. infrared waves
c. blue light
d. microwaves

18. Which form of EM radiation has the most penetrating ability?

a. red light
b. microwaves
c. gamma rays
d. infrared radiation

19. Why are high-frequency gamma rays more dangerous to humans than visible light?

a. Gamma rays have a lower frequency range than

visible light.
b. Gamma rays have a longer wavelength range than visible light.
c. Gamma rays have greater energy than visible light for penetrating matter.
d. Gamma rays have less energy than visible light for penetrating matter.

20. A dog would have a hard time stalking and catching a red bird hiding in a field of green grass. Explain this in terms of cone cells and color perception.
a. Dogs are red-green color-blind because they can see only blue and yellow through two kinds of cone cells present in their eyes.
b. Dogs are only red color-blind because they can see only blue and yellow through two kinds of cones cells present in their eyes.
c. Dogs are only green color-blind because they can see only blue and yellow through two kinds of cones cells present in their eyes.
d. Dogs are color-blind because they have only rods and no cone cells present in their eyes.

15.2 The Behavior of Electromagnetic Radiation

21. To compare the brightness of light bulbs for sale in a store, you should look on the labels to see how they are rated in terms of ____.
a. frequency
b. watts
c. amps
d. lumens

22. What is the wavelength of red light with a frequency of 4.00×10^{14} Hz?
a. 2.50×10^{14} m
b. 4.00×10^{15} m
c. 2.50×10^{6} m
d. 4.00×10^{-7} m

23. What is the distance of one light year in kilometers?
a. 2.59×10^{10} km
b. 1.58×10^{11} km
c. 2.63×10^{9} km
d. 9.46×10^{12} km

24. How does the illuminance of light change when the distance from the light source is tripled? Cite the relevant equation and explain how it supports your answer.
a. $\text{Illuminance} = \frac{P}{4\pi r^2}$; if distance is tripled, then the illuminance increases by 19 times.
b. $\text{Illuminance} = \frac{P}{4\pi r}$; if distance is tripled, then the illuminance decreases by 13 times.
c. $\text{Illuminance} = P \cdot 4\pi r^2$; if distance is tripled, then the illuminance decreases by 9 times.
d. $\text{Illuminance} = P \cdot 4\pi r$; if distance tripled, then the illuminance increases by 3 times.

25. A light bulb has an illuminance of 19.9 lx at a distance of 2 m . What is the luminous flux of the bulb?
a. 500 lm
b. 320 lm
c. 250 lm
d. 1,000 lm

Short Answer

15.1 The Electromagnetic Spectrum

26. Describe one way in which heat waves—infrared radiation—are different from sound waves.
a. Sound waves are transverse waves, whereas heat waves—infrared radiation—are longitudinal waves.
b. Sound waves have shorter wavelengths than heat waves.
c. Sound waves require a medium, whereas heat waves—infrared radiation—do not.
d. Sound waves have higher frequencies than heat waves.

27. Describe the electric and magnetic fields that make up an electromagnetic wave in terms of their orientation relative to each other and their phases.
a. They are perpendicular to and out of phase with each other.
b. They are perpendicular to and in phase with each other.
c. They are parallel to and out of phase with each other.
d. They are parallel to and in phase with each other.

28. Explain how X-radiation can be harmful and how it can be a useful diagnostic tool.
a. Overexposure to X-rays can cause HIV, though normal levels of X-rays can be used for sterilizing needles.
b. Overexposure to X-rays can cause cancer, though in limited doses X-rays can be used for imaging internal body parts.
c. Overexposure to X-rays causes diabetes, though normal levels of X-rays can be used for imaging internal body parts.
d. Overexposure to X-rays causes cancer, though normal levels of X-rays can be used for reducing

cholesterol in the blood.

29. Explain how sunlight is the original source of the energy in the food we eat.
   a. Sunlight is converted into chemical energy by plants; this energy is released when we digest food.
   b. Sunlight is converted into chemical energy by animals; this energy is released when we digest food.
   c. Sunlight is converted into chemical energy by fish; this energy is released when we digest food.
   d. Sunlight is converted into chemical energy by humans; this energy is released when we digest food.

15.2 The Behavior of Electromagnetic Radiation

30. Describe what happens to the path of light when the light slows down as it passes from one medium to another?
   a. The path of the light remains the same.
   b. The path of the light becomes circular.
   c. The path of the light becomes curved.
   d. The path of the light changes.

31. What is it about the nature of light reflected from snow that causes skiers to wear polarized sunglasses?
   a. The reflected light is polarized in the vertical direction.
   b. The reflected light is polarized in the horizontal direction.
   c. The reflected light has less intensity than the incident light.
   d. The reflected light has triple the intensity of the incident light.

32. How many lumens are radiated from a candle which has an illuminance of 3.98 lx at a distance of 2.00 m?
   a. 400 lm
   b. 100 lm
   c. 50 lm
   d. 200 lm

33. Saturn is 1.43×1012 m from the Sun. How many minutes does it take the Sun's light to reach Saturn?
   a. 7.94×10^{9} minutes
   b. 3.4×10^{4} minutes
   c. 3.4×10^{-6} minutes
   d. 79.4 minutes

Extended Response

15.1 The Electromagnetic Spectrum

34. A frequency of red light has a wavelength of 700 nm.
Part A—Compare the wavelength and frequency of violet light to red light.
Part B—Identify a type of radiation that has lower frequencies than red light.
Part C—Identify a type of radiation that has shorter wavelengths than violet light.
   a. A. Violet light has a lower frequency and longer wavelength than red light.
      B. ultraviolet radiation
      C. infrared radiation
   b. A. Violet light has a lower frequency and longer wavelength than red light.
      B. infrared radiation
      C. ultraviolet radiation
   c. A. Violet light has a higher frequency and shorter wavelength than red light.
      B. ultraviolet radiation
      C. infrared radiation
   d. A. Violet light has a higher frequency and shorter wavelength than red light.
      B. infrared radiation
      C. ultraviolet radiation

35. A mixture of red and green light is shone on each of the subtractive colors.
Part A—Which of these colors of light are reflected from magenta?
Part B—Which of these colors of light are reflected from yellow?
Part C—Which these colors of light are reflected from cyan?
   a. Part A. red and green
      Part B. green
      Part C. red
   b. Part A. red and green
      Part B. red
      Part C. green
   c. Part A. green
      Part B. red and green
      Part C. red
   d. Part A. red
      Part B. red and green
      Part C. green

15.2 The Behavior of Electromagnetic Radiation

36. Explain why we see the colorful effects of thin-film interference on the surface of soap bubbles and oil

slicks, but not on the surface of a window pane or clear plastic bag.

a. The thickness of a window pane or plastic bag is more than the wavelength of light, and interference occurs for thicknesses smaller than the wavelength of light.
b. The thickness of a window pane or plastic bag is less than the wavelength of light, and interference occurs for thicknesses similar to the wavelength of light.
c. The thickness of a window pane or plastic bag is more than the wavelength of light, and interference occurs for thicknesses similar to the wavelength of light.
d. The thickness of a window pane or plastic bag is less than the wavelength of light, and interference occurs for thicknesses larger than the wavelength of light.

37. The Occupational Safety and Health Administration (OSHA) recommends an illuminance of $500\ \mathrm{lx}$ for desktop lighting. An office space has lighting hung $2.50\ \mathrm{m}$ above desktop level that provides only $300\ \mathrm{lx}$. To what height would the lighting fixtures have to be lowered to provide $500\ \mathrm{lx}$ on desktops?

a. $1.22\ \mathrm{m}$
b. $1.09\ \mathrm{m}$
c. $0.96\ \mathrm{m}$
d. $1.94\ \mathrm{m}$

CHAPTER 16
Mirrors and Lenses

Figure 16.1 Flat, smooth surfaces reflect light to form mirror images. (credit: NASA Goddard Photo and Video, via Flickr)

Chapter Outline

16.1 Reflection

16.2 Refraction

16.3 Lenses

INTRODUCTION "In another moment Alice was through the glass, and had jumped lightly down into the Looking-glass room."

—Through the Looking Glass *by Lewis Carol*

Through the Looking Glass tells of the adventures of Alice after she steps from the real world, through a mirror, and into the virtual world. In this chapter we examine the optical meanings of real and virtual, as well as other concepts that make up the field of optics.

The light from this page or screen is formed into an image by the lens of your eyes, much as the lens of the camera that made the photograph at the beginning of this chapter. Mirrors, like lenses, can also form images, which in turn are captured by your eyes.

Optics is the branch of physics that deals with the behavior of visible light and other electromagnetic waves. For now, we concentrate on the propagation of light and its interaction with matter.

It is convenient to divide optics into two major parts based on the size of objects that light encounters. When light interacts with an object that is several times as large as the light's wavelength, its observable behavior is similar to a ray; it does not display its

wave characteristics prominently. We call this part of optics *geometric optics*. This chapter focuses on situations for which geometric optics is suited.

16.1 Reflection

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Explain reflection from mirrors, describe image formation as a consequence of reflection from mirrors, apply ray diagrams to predict and interpret image and object locations, and describe applications of mirrors
- Perform calculations based on the law of reflection and the equations for curved mirrors

Section Key Terms

angle of incidence	angle of reflection	central axis	concave mirror	convex mirror
diffused	focal length	focal point	geometric optics	law of reflection
law of refraction	ray	real image	specular	virtual image

Characteristics of Mirrors

There are three ways, as shown in Figure 16.2, in which light can travel from a source to another location. It can come directly from the source through empty space, such as from the Sun to Earth. Light can travel to an object through various media, such as air and glass. Light can also arrive at an object after being reflected, such as by a mirror. In all these cases, light is modeled as traveling in a straight line, called a ***ray***. Light may change direction when it encounters the surface of a different material (such as a mirror) or when it passes from one material to another (such as when passing from air into glass). It then continues in a straight line—that is, as a ray. The word *ray* comes from mathematics. Here it means a straight line that originates from some point. It is acceptable to visualize light rays as laser rays (or even science fiction depictions of *ray guns*).

Figure 16.2 Three methods for light to travel from a source to another location are shown. (a) Light reaches the upper atmosphere of Earth by traveling through empty space directly from the source (the Sun). (b) This light can reach a person in one of two ways. It can travel through a medium, such as air or glass, and typically travels from one medium to another. It can also reflect from an object, such as a mirror.

Because light moves in straight lines, that is, as rays, and changes directions when it interacts with matter, it can be described through geometry and trigonometry. This part of optics, described by straight lines and angles, is therefore called ***geometric optics***. There are two laws that govern how light changes direction when it interacts with matter: the **law of reflection**, for situations in which light bounces off matter; and the **law of refraction**, for situations in which light passes through matter. In this section, we consider the geometric optics of **reflection**.

Whenever we look into a mirror or squint at sunlight glinting from a lake, we are seeing a reflection. How does the reflected light travel from the object to your eyes? The law of reflection states: The **angle of reflection**, θ_r, equals the **angle of incidence**, θ_i

. This law governs the behavior of all waves when they interact with a smooth surface, and therefore describe the behavior of light waves as well. The reflection of light is simplified when light is treated as a ray. This concept is illustrated in Figure 16.3, which also shows how the angles are measured relative to the line perpendicular to the surface at the point where the light ray strikes it. This perpendicular line is also called the *normal line*, or just the *normal*. Light reflected in this way is referred to as **specular** (from the Latin word for *mirror*: *speculum*).

We expect to see reflections from smooth surfaces, but Figure 16.4, illustrates how a rough surface reflects light. Because the light is reflected from different parts of the surface at different angles, the rays go in many different directions, so the reflected light is **diffused**. Diffused light allows you to read a printed page from almost any angle because some of the rays go in different directions. Many objects, such as people, clothing, leaves, and walls, have rough surfaces and can be seen from many angles. A mirror, on the other hand, has a smooth surface and reflects light at specific angles.

Figure 16.3 The law of reflection states that the angle of reflection, $\boldsymbol{\theta_r}$, equals the angle of incidence, $\boldsymbol{\theta_i}$. The angles are measured relative to the line perpendicular to the surface at the point where the ray strikes the surface. The incident and reflected rays, along with the normal, lie in the same plane.

Figure 16.4 Light is diffused when it reflects from a rough surface. Here, many parallel rays are incident, but they are reflected at many different angles because the surface is rough.

When we see ourselves in a mirror, it appears that our image is actually behind the mirror. We see the light coming from a direction determined by the law of reflection. The angles are such that our image is exactly the same distance behind the mirror, d_i, as the distance we stand away from the mirror, d_o. Although these mirror images make objects appear to be where they cannot be (such as behind a solid wall), the images are not figments of our imagination. Mirror images can be photographed and videotaped by instruments and look just as they do to our eyes, which are themselves optical instruments. An image in a mirror is said to be a **virtual image**, as opposed to a **real image**. A virtual image is formed when light rays appear to diverge from a point without actually doing so.

Figure 16.5 helps illustrate how a flat mirror forms an image. Two rays are shown emerging from the same point, striking the mirror, and reflecting into the observer's eye. The rays can diverge slightly, and both still enter the eye. If the rays are extrapolated backward, they seem to originate from a common point behind the mirror, allowing us to locate the image. The paths of the reflected rays into the eye are the same as if they had come directly from that point behind the mirror. Using the law of reflection—the angle of reflection equals the angle of incidence—we can see that the image and object are the same distance from the mirror. This is a virtual image, as defined earlier.

Figure 16.5 When two sets of rays from common points on an object are reflected by a flat mirror into the eye of an observer, the reflected rays seem to originate from behind the mirror, which determines the position of the virtual image.

FUN IN PHYSICS

Mirror Mazes

Figure 16.6 is a chase scene from an old silent film called *The Circus*, starring Charlie Chaplin. The chase scene takes place in a mirror maze. You may have seen such a maze at an amusement park or carnival. Finding your way through the maze can be very difficult. Keep in mind that only one image in the picture is real—the others are virtual.

Figure 16.6 Charlie Chaplin is in a mirror maze. Which image is real?

One of the earliest uses of mirrors for creating the illusion of space is seen in the Palace of Versailles, the former home of French royalty. Construction of the Hall of Mirrors (Figure 16.7) began in 1678. It is still one of the most popular tourist attractions at Versailles.

Figure 16.7 Tourists love to wander in the Hall of Mirrors at the Palace of Versailles. (credit: Michal Osmenda, Flickr)

GRASP CHECK

Only one Charlie in this image (Figure 16.8) is real. The others are all virtual images of him. Can you tell which is real? Hint—His hat is tilted to one side.

Figure 16.8

a. The virtual images have their hats tilted to the right.
b. The virtual images have their hats tilted to the left.
c. The real images have their hats tilted to the right.
d. The real images have their hats tilted to the left.

WATCH PHYSICS

Virtual Image

This video explains the creation of virtual images in a mirror. It shows the location and orientation of the images using ray diagrams, and relates the perception to the human eye.

Click to view content (https://openstax.org/l/28Virtualimage)

Compare the distance of an object from a mirror to the apparent distance of its virtual image behind the mirror.

a. The distances of the image and the object from the mirror are the same.
b. The distances of the image and the object from the mirror are always different.
c. The image is formed at infinity if the object is placed near the mirror.
d. The image is formed near the mirror if the object is placed at infinity.

Some mirrors are curved instead of flat. A mirror that curves inward is called a ***concave mirror***, whereas one that curves outward is called a ***convex mirror***. Pick up a well-polished metal spoon and you can see an example of each type of curvature. The side of the spoon that holds the food is a concave mirror; the back of the spoon is a convex mirror. Observe your image on both sides of the spoon.

TIPS FOR SUCCESS

You can remember the difference between concave and convex by thinking, *Concave means caved in*.

Ray diagrams can be used to find the point where reflected rays converge or appear to converge, or the point from which rays appear to diverge. This is called the ***focal point***, F. The distance from F to the mirror along the **central axis** (the line perpendicular to the center of the mirror's surface) is called the ***focal length***, *f*. Figure 16.9 shows the focal points of concave and convex mirrors.

Figure 16.9 (a, b) The focal length for the concave mirror in (a), formed by converging rays, is in front of the mirror, and has a positive value. The focal length for the convex mirror in (b), formed by diverging rays, appears to be behind the mirror, and has a negative value.

Images formed by a concave mirror vary, depending on which side of the focal point the object is placed. For any object placed on the far side of the focal point with respect to the mirror, the rays converge in front of the mirror to form a real image, which can be projected onto a surface, such as a screen or sheet of paper However, for an object located inside the focal point with respect to the concave mirror, the image is virtual. For a convex mirror the image is always virtual—that is, it appears to be behind the mirror. The ray diagrams in Figure 16.10 show how to determine the nature of the image formed by concave and convex mirrors.

Figure 16.10 (a) The image of an object placed outside the focal point of a concave mirror is inverted and real. (b) The image of an object

placed inside the focal point of a concave mirror is erect and virtual. (c) The image of an object formed by a convex mirror is erect and virtual.

The information in Figure 16.10 is summarized in Table 16.1.

Type of Mirror	Object to Mirror Distance, d_o	Image Characteristics
Concave	$d_o > f$	Real and inverted
Concave	$d_o < f$	Virtual and erect
Convex	$d_o <$ or $> f$	Virtual and erect

Table 16.1 Curved Mirror Images This table details the type and orientation of images formed by concave and convex mirrors.

Snap Lab

Concave and Convex Mirrors

- Silver spoon and silver polish, or a new spoon made of any shiny metal

Instructions

Procedure

1. Choose any small object with a top and a bottom, such as a short nail or tack, or a coin, such as a quarter. Observe the object's reflection on the back of the spoon.
2. Observe the reflection of the object on the front (bowl side) of the spoon when held away from the spoon at a distance of several inches.
3. Observe the image while slowly moving the small object toward the bowl of the spoon. Continue until the object is all the way inside the bowl of the spoon.
4. You should see one point where the object disappears and then reappears. This is the focal point.

WATCH PHYSICS

Parabolic Mirrors and Real Images

This video uses ray diagrams to show the special feature of parabolic mirrors that makes them ideal for either projecting light energy in parallel rays, with the source being at the focal point of the parabola, or for collecting at the focal point light energy from a distant source.

Click to view content (https://www.openstax.org/l/28Parabolic)

Explain why using a parabolic mirror for a car headlight throws much more light on the highway than a flat mirror.

a. The rays do not polarize after reflection.
b. The rays are dispersed after reflection.
c. The rays are polarized after reflection.
d. The rays become parallel after reflection.

You should be able to notice everyday applications of curved mirrors. One common example is the use of security mirrors in stores, as shown in Figure 16.11.

Figure 16.11 Security mirrors are convex, producing a smaller, upright image. Because the image is smaller, a larger area is imaged compared with what would be observed for a flat mirror; hence, security is improved. (credit: Laura D'Alessandro, Flickr)

Some telescopes also use curved mirrors and no lenses (except in the eyepieces) both to magnify images and to change the path of light. Figure 16.12 shows a Schmidt-Cassegrain telescope. This design uses a spherical primary concave mirror and a convex secondary mirror. The image is projected onto the focal plane by light passing through the perforated primary mirror. The effective focal length of such a telescope is the focal length of the primary mirror multiplied by the magnification of the secondary mirror. The result is a telescope with a focal length much greater than the length of the telescope itself.

Figure 16.12 This diagram shows the design of a Schmidt–Cassegrain telescope.

A parabolic concave mirror has the very useful property that all light from a distant source, on reflection by the mirror surface, is directed to the focal point. Likewise, a light source placed at the focal point directs all the light it emits in parallel lines away from the mirror. This case is illustrated by the ray diagram in Figure 16.13. The light source in a car headlight, for example, is located at the focal point of a parabolic mirror.

Figure 16.13 The bulb in this ray diagram of a car headlight is located at the focal point of a parabolic mirror.

Parabolic mirrors are also used to collect sunlight and direct it to a focal point, where it is transformed into heat, which in turn can be used to generate electricity. This application is shown in Figure 16.14.

Figure 16.14 Parabolic trough collectors are used to generate electricity in southern California. (credit: kjkolb, Wikimedia Commons)

Using a concave mirror, you look at the reflection of a faraway object. The image size changes if you move the object closer to the mirror. Why does the image disappear entirely when the object is at the mirror's focal point?

a. The height of the image became infinite.
b. The height of the object became zero.
c. The intensity of intersecting light rays became zero.
d. The intensity of intersecting light rays increased.

The Application of the Curved Mirror Equations

Curved mirrors and the images they create involve a fairly small number of variables: the mirror's radius of curvature, R; the focal length, f; the distances of the object and image from the mirror, d_o and d_i, respectively; and the heights of the object and image, h_o and h_i, respectively. The signs of these values indicate whether the image is inverted, erect (upright), real, or virtual. We now look at the equations that relate these variables and apply them to everyday problems.

Figure 16.15 shows the meanings of most of the variables we will use for calculations involving curved mirrors.

Figure 16.15 Look for the variables, d_o, d_i, h_o, h_i, and f in this figure.

The basic equation that describes both lenses and mirrors is the lens/mirror equation

$$\frac{1}{f} = \frac{1}{d_i} + \frac{1}{d_o}.$$

This equation can be rearranged several ways. For example, it may be written to solve for focal length.

$$f = \frac{d_i d_o}{d_o + d_i}$$

Magnification, m, is the ratio of the size of the image, h_i, to the size of the object, h_o. The value of m can be calculated in two ways.

$$m = \frac{h_i}{h_o} = \frac{-d_i}{d_o}$$

This relationship can be written to solve for any of the variables involved. For example, the height of the image is given by

$$h_i = -h_o\left(\frac{d_i}{d_o}\right).$$

We saved the simplest equation for last. The radius of curvature of a curved mirror, R, is simply twice the focal length.

$$R = 2f$$

We can learn important information from the algebraic sign of the result of a calculation using the previous equations:

- A negative d_i indicates a virtual image; a positive value indicates a real image
- A negative h_i indicates an inverted image; a positive value indicates an erect image
- For concave mirrors, f is positive; for convex mirrors, f is negative

Now let's apply these equations to solve some problems.

WORKED EXAMPLE

Calculating Focal Length

A person standing 6.0 m from a convex security mirror forms a virtual image that appears to be 1.0 m behind the mirror. What is the focal length of the mirror?

STRATEGY

The person is the object, so $d_o = 6.0$ m. We know that, for this situation, d_o is positive. The image is virtual, so the value for the image distance is negative, so $d_i = -1.0$ m.

Now, use the appropriate version of the lens/mirror equation to solve for focal length by substituting the known values.

Solution

$$f = \frac{d_i d_o}{d_o + d_i} = \frac{(-1.0)(6.0)}{6.0 + (-1.0)} = \frac{-6.0}{5.0} = -1.2 \text{ m}$$

Discussion

The negative result is expected for a convex mirror. This indicates the focal point is behind the mirror.

WORKED EXAMPLE

Calculating Object Distance

Electric room heaters use a concave mirror to reflect infrared (IR) radiation from hot coils. Note that IR radiation follows the same law of reflection as visible light. Given that the mirror has a radius of curvature of 50.0 cm and produces an image of the coils 3.00 m in front of the mirror, where are the coils with respect to the mirror?

STRATEGY

We are told that the concave mirror projects a real image of the coils at an image distance $d_i = 3.00$ m. The coils are the object, and we are asked to find their location—that is, to find the object distance d_o. We are also given the radius of curvature of the mirror, so that its focal length is $f = R/2 = 25.0$ cm (a positive value, because the mirror is concave, or converging). We can use the lens/mirror equation to solve this problem.

Solution

Because d_i and f are known, the lens/mirror equation can be used to find d_o.

$$\frac{1}{f} = \frac{1}{d_i} + \frac{1}{d_o} \tag{16.1}$$

Rearranging to solve for d_o, we have

$$d_o = \frac{d_i f}{d_i - f}. \tag{16.2}$$

Entering the known quantities gives us

$$d_o = \frac{(3.00 \times 10^2)(25.0)}{(3.00 \times 10^2) - 25.0} = 27.3 \text{ cm}. \tag{16.3}$$

Discussion

Note that the object (the coil filament) is farther from the mirror than the mirror's focal length. This is a *case 1* image ($d_o > f$ and f positive), consistent with the fact that a real image is formed. You get the most concentrated thermal energy directly in front of the mirror and 3.00 m away from it. In general, this is not desirable because it could cause burns. Usually, you want the rays to emerge parallel, and this is accomplished by having the filament at the focal point of the mirror.

Note that the filament here is not much farther from the mirror than the focal length, and that the image produced is considerably farther away.

Practice Problems

1. A concave mirror has a radius of curvature of $0.8\ \text{m}$. What is the focal length of the mirror?
   a. $-0.8\ \text{m}$
   b. $-0.4\ \text{m}$
   c. $0.4\ \text{m}$
   d. $0.8\ \text{m}$

2. What is the focal length of a makeup mirror that produces a magnification of 1.50 when a person's face is 12.0 cm away? Construct a ray diagram using paper, a pencil and a ruler to confirm your calculation.
   a. −36.0 cm
   b. −7.20 cm
   c. 7.20 cm
   d. 36.0 cm

Check Your Understanding

3. How does the object distance, d_o, compare with the focal length, f, for a concave mirror that produces an image that is real and inverted?
   a. $d_o > f$, where d_o and f are object distance and focal length, respectively.
   b. $d_o < f$, where d_o and f are object distance and focal length, respectively.
   c. $d_o = f$, where do and f are object distance and focal length, respectively.
   d. $d_o = 0$, where do is the object distance.

4. Use the law of reflection to explain why it is not a good idea to polish a mirror with sandpaper.
   a. The surface becomes smooth, and a smooth surface produces a sharp image.
   b. The surface becomes irregular, and an irregular surface produces a sharp image.
   c. The surface becomes smooth, and a smooth surface transmits light, but does not reflect it.
   d. The surface becomes irregular, and an irregular surface produces a blurred image.

5. An object is placed in front of a concave mirror at a distance that is greater than the focal length of the mirror. Will the image produced by the mirror be real or virtual? Will it be erect or inverted?
   a. It is real and erect.
   b. It is real and inverted.
   c. It is virtual and inverted.
   d. It is virtual and erect.

16.2 Refraction

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Explain refraction at media boundaries, predict the path of light after passing through a boundary (Snell's law), describe the index of refraction of materials, explain total internal reflection, and describe applications of refraction and total internal reflection
- Perform calculations based on the law of refraction, Snell's law, and the conditions for total internal reflection

Section Key Terms

angle of refraction	corner reflector	critical angle	dispersion	incident ray
index of refraction	refracted ray	Snell's law	total internal reflection	

The Law of Refraction

You may have noticed some odd optical phenomena when looking into a fish tank. For example, you may see the same fish appear to be in two different places (Figure 16.16). This is because light coming to you from the fish changes direction when it

leaves the tank and, in this case, light rays traveling along two different paths both reach our eyes. The changing of a light ray's direction (loosely called *bending*) when it passes a boundary between materials of different composition, or between layers in single material where there are changes in temperature and density, is called *refraction*. Refraction is responsible for a tremendous range of optical phenomena, from the action of lenses to voice transmission through optical fibers.

Figure 16.16 Looking at the fish tank as shown, we can see the same fish in two different locations, because light changes directions when it passes from water to air. In this case, light rays traveling on two different paths change direction as they travel from water to air, and so reach the observer. Consequently, the fish appears to be in two different places. This bending of light is called *refraction* and is responsible for many optical phenomena.

Why does light change direction when passing from one material (medium) to another? It is because light changes speed when going from one material to another. This behavior is typical of all waves and is especially easy to apply to light because light waves have very small wavelengths, and so they can be treated as rays. Before we study the law of refraction, it is useful to discuss the speed of light and how it varies between different media.

The speed of light is now known to great precision. In fact, the speed of light in a vacuum, c, is so important, and is so precisely known, that it is accepted as one of the basic physical quantities, and has the fixed value

$$c = 2.9972458 \times 10^8 \text{m/s} \approx 3.00 \times 10^8 \text{m/s} \tag{16.4}$$

where the approximate value of 3.00×10^8 m/s is used whenever three-digit precision is sufficient. The speed of light through matter is less than it is in a vacuum, because light interacts with atoms in a material. The speed of light depends strongly on the type of material, given that its interaction with different atoms, crystal lattices, and other substructures varies. We define the **index of refraction**, n, of a material to be

$$n = \frac{c}{v},$$

where v is the observed speed of light in the material. Because the speed of light is always less than c in matter and equals c only in a vacuum, the index of refraction (plural: indices of refraction) is always greater than or equal to one.

Table 16.2 lists the indices of refraction in various common materials.

Medium	n
Gases at 0 °C and 1 atm	

Table 16.2 Indices of Refraction The table lists the indices of refraction for various materials that are transparent to light. Note, that light travels the slowest in the materials with the greatest indices of refraction.

Medium	*n*
Air	1.000293
Carbon dioxide	1.00045
Hydrogen	1.000139
Oxygen	1.000271
Liquids at 20 °C	
Benzene	1.501
Carbon disulfide	1.628
Carbon tetrachloride	1.461
Ethanol	1.361
Glycerin	1.473
Water, fresh	1.333
Solids at 20 °C	
Diamond	2.419
Fluorite	1.434
Glass, crown	1.52
Glass, flint	1.66
Ice at 0 °C	1.309
Plexiglas	1.51
Polystyrene	1.49
Quartz, crystalline	1.544
Quartz, fused	1.458
Sodium chloride	1.544

Table 16.2 Indices of Refraction The table lists the indices of refraction for various materials that are transparent to light. Note, that light travels the slowest in the materials with the greatest indices of refraction.

Medium	*n*
Zircon	1.923

Table 16.2 Indices of Refraction The table lists the indices of refraction for various materials that are transparent to light. Note, that light travels the slowest in the materials with the greatest indices of refraction.

Figure 16.17 provides an analogy for and a description of how a ray of light changes direction when it passes from one medium to another. As in the previous section, the angles are measured relative to a perpendicular to the surface at the point where the light ray crosses it. The change in direction of the light ray depends on how the speed of light changes. The change in the speed of light is related to the indices of refraction of the media involved. In the situations shown in Figure 16.17, medium 2 has a greater index of refraction than medium 1. This difference in index of refraction means that the speed of light is less in medium 2 than in medium 1. Note that, in Figure 16.17(a), the path of the ray moves closer to the perpendicular when the ray slows down. Conversely, in Figure 16.17(b), the path of the ray moves away from the perpendicular when the ray speeds up. The path is exactly reversible. In both cases, you can imagine what happens by thinking about pushing a lawn mower from a footpath onto grass, and vice versa. Going from the footpath to grass, the right front wheel is slowed and pulled to the side as shown. This is the same change in direction for light when it goes from a fast medium to a slow one. When going from the grass to the footpath, the left front wheel moves faster than the others, and the mower changes direction as shown. This, too, is the same change in direction as light going from slow to fast.

Figure 16.17 The change in direction of a light ray depends on how the speed of light changes when it crosses from one medium to another. For the situations shown here, the speed of light is greater in medium 1 than in medium 2. (a) A ray of light moves closer to the perpendicular when it slows down. This is analogous to what happens when a lawnmower goes from a footpath (medium 1) to grass (medium 2). (b) A ray of light moves away from the perpendicular when it speeds up. This is analogous to what happens when a lawnmower goes from grass (medium 2) to the footpath (medium 1). The paths are exactly reversible.

Snap Lab

Bent Pencil

A classic observation of refraction occurs when a pencil is placed in a glass filled halfway with water. Do this and observe the shape of the pencil when you look at it sideways through air, glass, and water.

- A full-length pencil
- A glass half full of water

Instructions

Procedure

1. Place the pencil in the glass of water.
2. Observe the pencil from the side.

3. Explain your observations.

Virtual Physics

Bending Light

Click to view content (https://www.openstax.org/l/28Bendinglight)

The Bending Light simulation in allows you to show light refracting as it crosses the boundaries between various media (download animation first to view). It also shows the reflected ray. You can move the protractor to the point where the light meets the boundary and measure the angle of incidence, the **angle of refraction**, and the angle of reflection. You can also insert a prism into the beam to view the spreading, or **dispersion**, of white light into colors, as discussed later in this section. Use the ray option at the upper left.

A light ray moving upward strikes a horizontal boundary at an acute angle relative to the perpendicular and enters the medium above the boundary. What must be true for the light to bend away from the perpendicular?

a. The medium below the boundary must have a greater index of refraction than the medium above.
b. The medium below the boundary must have a lower index of refraction than the medium above.
c. The medium below the boundary must have an index of refraction of zero.
d. The medium above the boundary must have an infinite index of refraction.

The amount that a light ray changes direction depends both on the incident angle and the amount that the speed changes. For a ray at a given incident angle, a large change in speed causes a large change in direction, and thus a large change in the angle of refraction. The exact mathematical relationship is the law of refraction, or ***Snell's law***, which is stated in equation form as

$$n_1 \sin\theta_1 = n_2 \sin\theta_2 \text{ or } \frac{n_1}{n_2} = \frac{\sin\theta_2}{\sin\theta_1}.$$

In terms of speeds, Snell's law becomes

$$\frac{\sin\theta_1}{\sin\theta_2} = \frac{v_1}{v_2}. \tag{16.5}$$

Here, n_1 and n_2 are the indices of refraction for media 1 and 2, respectively, and θ_1 and θ_2 are the angles between the rays and the perpendicular in the respective media 1 and 2, as shown in Figure 16.17. The incoming ray is called the ***incident ray*** and the outgoing ray is called the ***refracted ray***. The associated angles are called the *angle of incidence* and the *angle of refraction*. Later, we apply Snell's law to some practical situations.

Dispersion is defined as the spreading of white light into the wavelengths of which it is composed. This happens because the index of refraction varies slightly with wavelength. Figure 16.18 shows how a prism disperses white light into the colors of the rainbow.

Figure 16.18 (a) A pure wavelength of light (λ) falls onto a prism and is refracted at both surfaces. (b) White light is dispersed by the prism (spread of light exaggerated). Because the index of refraction varies with wavelength, the angles of refraction vary with wavelength. A sequence of red to violet is produced, because the index of refraction increases steadily with decreasing wavelength.

Rainbows are produced by a combination of refraction and reflection. You may have noticed that you see a rainbow only when you turn your back to the Sun. Light enters a drop of water and is reflected from the back of the drop, as shown in Figure 16.19. The light is refracted both as it enters and as it leaves the drop. Because the index of refraction of water varies with wavelength, the light is dispersed and a rainbow is observed.

Figure 16.19 Part of the light falling on this water drop enters and is reflected from the back of the drop. This light is refracted and dispersed both as it enters and as it leaves the drop.

WATCH PHYSICS

Dispersion

This video explains how refraction disperses white light into its composite colors.

Click to view content (https://www.openstax.org/l/28Raindrop)

Which colors of the rainbow bend most when refracted?

a. Colors with a longer wavelength and higher frequency bend most when refracted.
b. Colors with a shorter wavelength and higher frequency bend most when refracted.
c. Colors with a shorter wavelength and lower frequency bend most when refracted.
d. Colors with a longer wavelength and a lower frequency bend most when refracted.

A good-quality mirror reflects more than 90 percent of the light that falls on it; the mirror absorbs the rest. But, it would be useful to have a mirror that reflects all the light that falls on it. Interestingly, we can produce total reflection using an aspect of refraction. Consider what happens when a ray of light strikes the surface between two materials, such as is shown in Figure 16.20(a). Part of the light crosses the boundary and is refracted; the rest is reflected. If, as shown in the figure, the index of refraction for the second medium is less than the first, the ray bends away from the perpendicular. Because $n_1 > n_2$, the angle of refraction is greater than the angle of incidence—that is, $\theta_2 > \theta_1$. Now, imagine what happens as the incident angle is increased. This causes θ_2 to increase as well. The largest the angle of refraction, θ_2, can be is 90°, as shown in Figure 16.20(b). The **critical angle**, θ_c, for a combination of two materials is defined to be the incident angle, θ_1, which produces an angle of refraction of 90°. That is, θ_c is the incident angle for which $\theta_2 = 90°$. If the incident angle, θ_1, is greater than the critical angle, as shown in Figure 16.20(c), then all the light is reflected back into medium 1, a condition called ***total internal reflection***.

Figure 16.20 (a) A ray of light crosses a boundary where the speed of light increases and the index of refraction decreases—that is, $n_2 < n_1$. The refracted ray bends away from the perpendicular. (b) The critical angle, θ_c, is the one for which the angle of refraction is 90°. (c) Total internal reflection occurs when the incident angle is greater than the critical angle.

Recall that Snell's law states the relationship between angles and indices of refraction. It is given by

$$n_1 \sin\theta_1 = n_2 \sin\theta_2. \tag{16.6}$$

When the incident angle equals the critical angle ($\theta_1 = \theta_c$), the angle of refraction is 90° ($\theta_2 = 90°$). Noting that sin 90° = 1, Snell's law in this case becomes

$$n_1 \sin\theta_1 = n_2.$$

The critical angle, θ_c, for a given combination of materials is thus

$$\theta_c = \sin^{-1}\left(\frac{n_2}{n_1}\right),$$

for $n_1 > n_2$.

Total internal reflection occurs for any incident angle greater than the critical angle, θ_c, and it can only occur when the second medium has an index of refraction less than the first. Note that the previous equation is written for a light ray that travels in medium 1 and reflects from medium 2, as shown in Figure 16.20.

There are several important applications of total internal reflection. Total internal reflection, coupled with a large index of refraction, explains why diamonds sparkle more than other materials. The critical angle for a diamond-to-air surface is only 24.4°; so, when light enters a diamond, it has trouble getting back out (Figure 16.21). Although light freely enters the diamond at different angles, it can exit only if it makes an angle less than 24.4° with the normal to a given surface. Facets on diamonds are specifically intended to make this unlikely, so that the light can exit only in certain places. Diamonds with very few impurities are very clear, so the light makes many internal reflections and is concentrated at the few places it can exit—hence the sparkle.

Figure 16.21 Light cannot escape a diamond easily because its critical angle with air is so small. Most reflections are total and the facets are placed so that light can exit only in particular ways, thus concentrating the light and making the diamond sparkle.

A light ray that strikes an object that consists of two mutually perpendicular reflecting surfaces is reflected back exactly parallel to the direction from which it came. This parallel reflection is true whenever the reflecting surfaces are perpendicular, and it is independent of the angle of incidence. Such an object is called a ***corner reflector*** because the light bounces from its inside corner. Many inexpensive reflector buttons on bicycles, cars, and warning signs have corner reflectors designed to return light in the direction from which it originates. Corner reflectors are perfectly efficient when the conditions for total internal reflection are satisfied. With common materials, it is easy to obtain a critical angle that is less than 45°. One use of these perfect *mirrors* is in binoculars, as shown in Figure 16.22. Another application is for periscopes used in submarines.

Figure 16.22 These binoculars use corner reflectors with total internal reflection to get light to the observer's eyes.

Fiber optics are one common application of total internal reflection. In communications, fiber optics are used to transmit telephone, internet, and cable TV signals, and they use the transmission of light down fibers of plastic or glass. Because the

fibers are thin, light entering one is likely to strike the inside surface at an angle greater than the critical angle and, thus, be totally reflected (Figure 16.23). The index of refraction outside the fiber must be smaller than inside, a condition that is satisfied easily by coating the outside of the fiber with a material that has an appropriate refractive index. In fact, most fibers have a varying refractive index to allow more light to be guided along the fiber through total internal reflection. Rays are reflected around corners as shown in the figure, making the fibers into tiny *light pipes*.

Figure 16.23 (a) Fibers in bundles are clad by a material that has a lower index of refraction than the core to ensure total internal reflection, even when fibers are in contact with one another. A single fiber with its cladding is shown. (b) Light entering a thin fiber may strike the inside surface at large, or grazing, angles, and is completely reflected if these angles exceed the critical angle. Such rays continue down the fiber, even following it around corners, because the angles of reflection and incidence remain large.

LINKS TO PHYSICS

Medicine: Endoscopes

A medical device called an *endoscope* is shown in Figure 16.24.

Figure 16.24 Endoscopes, such as the one drawn here, send light down a flexible fiber optic tube, which sends images back to a doctor in charge of performing a medical procedure.

The word *endoscope means looking inside*. Doctors use endoscopes to look inside hollow organs in the human body and inside body cavities. These devices are used to diagnose internal physical problems. Images may be transmitted to an eyepiece or sent to a video screen. Another channel is sometimes included to allow the use of small surgical instruments. Such surgical procedures include collecting biopsies for later testing, and removing polyps and other growths.

Identify the process that allows light and images to travel through a tube that is not straight.

a. The process is refraction of light.
b. The process is dispersion of light.
c. The process is total internal reflection of light.
d. The process is polarization of light.

Calculations with the Law of Refraction

The calculation problems that follow require application of the following equations:

$$n = \frac{c}{v}, \tag{16.8}$$

$$n_1 \sin\theta_1 = n_2 \sin\theta_2 \text{ or } \frac{n_1}{n_2} = \frac{\sin\theta_2}{\sin\theta_1},$$

and

$$\theta_c = \sin^{-1}\left(\frac{n_2}{n_1}\right), \text{ for } n_1 > n_2.$$

These are the equations for refractive index, the mathematical statement of the law of refraction (Snell's law), and the equation for the critical angle.

WATCH PHYSICS

Snell's Law Example 1

This video leads you through calculations based on the application of the equation that represents Snell's law.

Click to view content (https://www.openstax.org/l/28Snellslaw)

Which two types of variables are included in Snell's law?

a. The two types of variables are density of a material and the angle made by the light ray with the normal.
b. The two types of variables are density of a material and the thickness of a material.
c. The two types of variables are refractive index and thickness of each material.
d. The two types of variables are refractive index of a material and the angle made by a light ray with the normal.

WORKED EXAMPLE

Calculating Index of Refraction from Speed

Calculate the index of refraction for a solid medium in which the speed of light is 2.012×10^8 m/s, and identify the most likely substance, based on the previous table of indicies of refraction.

STRATEGY

We know the speed of light, c, is 3.00×10^8 m/s, and we are given v. We can simply plug these values into the equation for index of refraction, n.

Solution

$$n = \frac{c}{v} = \frac{3.00 \times 10^8 \text{m/s}}{2.012 \times 10^8 \text{m/s}} = 1.49 \tag{16.9}$$

This value matches that of polystyrene exactly, according to the table of indices of refraction (Table 16.2).

Discussion

The three-digit approximation for c is used, which in this case is all that is needed. Many values in the table are only given to three significant figures. Note that the units for speed cancel to yield a dimensionless answer, which is correct.

WORKED EXAMPLE

Calculating Index of Refraction from Angles

Suppose you have an unknown, clear solid substance immersed in water and you wish to identify it by finding its index of refraction. You arrange to have a beam of light enter it at an angle of 45.00°, and you observe the angle of refraction to be 40.30°. What are the index of refraction of the substance and its likely identity?

STRATEGY

We must use the mathematical expression for the law of refraction to solve this problem because we are given angle data, not speed data.

$$\frac{n_1}{n_2} = \frac{\sin\theta_2}{\sin\theta_1} \tag{16.10}$$

The subscripts 1 and 2 refer to values for water and the unknown, respectively, where 1 represents the medium from which the

light is coming and 2 is the new medium it is entering. We are given the angle values, and the table of indicies of refraction gives us n for water as 1.333. All we have to do before solving the problem is rearrange the equation

$$n_2 = \frac{n_1 \sin\theta_1}{\sin\theta_2}. \tag{16.11}$$

Solution

$$n_2 = \frac{(1.333)(0.7071)}{0.6468} = 1.457 \tag{16.12}$$

The best match from Table 16.2 is fused quartz, with $n = 1.458$.

Discussion

Note the relative sizes of the variables involved. For example, a larger angle has a larger sine value. This checks out for the two angles involved. Note that the smaller value of θ_2 compared with θ_1 indicates the ray has bent *toward* normal. This result is to be expected if the unknown substance has a greater n value than that of water. The result shows that this is the case.

WORKED EXAMPLE

Calculating Critical Angle

Verify that the critical angle for light going from water to air is 48.6°. (See Table 16.2, the table of indices of refraction.)

STRATEGY

First, choose the equation for critical angle

$$\theta_c = \sin^{-1}\left(\frac{n_2}{n_1}\right), \text{ for } n_1 > n_2. \tag{16.13}$$

Then, look up the n values for water, n_1, and air, n_2. Find the value of $\frac{n_2}{n_1}$. Last, find the angle that has a sine equal to this value and it compare with the given angle of 48.6°.

Solution

For water, $n_1 = 1.333$; for air, $n_2 = 1.0003$. So,

$$\begin{aligned} \frac{n_2}{n_1} &= \frac{1.0003}{1.333} = 0.7504 \\ \sin^{-1}(0.7504) &= 48.63°. \end{aligned} \tag{16.14}$$

Discussion

Remember, when we try to find a critical angle, we look for the angle at which light can no longer escape past a medium boundary by refraction. It is logical, then, to think of subscript 1 as referring to the medium the light is trying to leave, and subscript 2 as where it is trying (unsuccessfully) to go. So water is 1 and air is 2.

Practice Problems

6. The refractive index of ethanol is 1.36. What is the speed of light in ethanol?
   a. 2.25×108 m/s
   b. 2.21×107 m/s
   c. 2.25×109 m/s
   d. 2.21×108 m/s

7. The refractive index of air is 1.0003 and the refractive index of crystalline quartz is 1.544. What is the critical angle for a ray of light going from crystalline quartz into air?
   a. 49.61°
   b. 20.19°
   c. 0.6479 rad
   d. 0.7048 rad

Check Your Understanding

8. Which law is expressed by the equation $n_1 \sin \vartheta_1 = n_2 \sin \vartheta_2$?
   a. This is Ohm's law.
   b. This is Wien's displacement law.
   c. This is Snell's law.
   d. This is Newton's law.

9. Explain why the index of refraction is always greater than or equal to one.
   a. The formula for index of refraction, n, of a material is $n = \frac{\text{speed of light in a material}}{\text{speed of light in a vacuum}} = \frac{v}{c}$, where $\text{v} > \text{c}$, so n is always greater than one.
   b. The formula for index of refraction, n, of a material is $n = \frac{\text{speed of light in a vacuum}}{\text{speed of light in a material}} = \frac{c}{v}$, where $\text{c} > \text{v}$, so n is always greater than one.
   c. The formula for index of refraction, n, of a material is $n = \text{speed of light in a vacuum} \times \text{speed of light in a materaial} = c \times v$, where $\text{c, v} > 1$, so n is always greater than one.
   d. The formula for refractive index, n, of a material is $n = \frac{1}{\text{speed of light in a vacuum} \times \text{speed of light in a material}} = \frac{1}{c \times v}$, where $\text{c} < \text{v} < 1$, so n is always greater than one.

10. Write an equation that expresses the law of refraction.
   a. $\frac{n_1}{n_2} = \frac{\sin \theta_1}{\sin \theta_2}$
   b. $\frac{n_2}{n_1} = \left(\frac{\sin \theta_2}{\sin \theta_1}\right)^2$
   c. $\frac{n_1}{n_2} = \left(\frac{\sin \theta_2}{\sin \theta_1}\right)^2$
   d. $\frac{n_1}{n_2} = \frac{\sin \theta_2}{\sin \theta_1}$

16.3 Lenses

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Describe and predict image formation and magnification as a consequence of refraction through convex and concave lenses, use ray diagrams to confirm image formation, and discuss how these properties of lenses determine their applications
- Explain how the human eye works in terms of geometric optics
- Perform calculations, based on the thin-lens equation, to determine image and object distances, focal length, and image magnification, and use these calculations to confirm values determined from ray diagrams

Section Key Terms

aberration	chromatic aberration	concave lens	converging lens	convex lens
diverging lens	eyepiece	objective	ocular	parfocal

Characteristics of Lenses

Lenses are found in a huge array of optical instruments, ranging from a simple magnifying glass to the eye to a camera's zoom lens. In this section, we use the law of refraction to explore the properties of lenses and how they form images.

Some of what we learned in the earlier discussion of curved mirrors also applies to the study of lenses. Concave, convex, focal point F, and focal length *f* have the same meanings as before, except each measurement is made from the center of the lens instead of the surface of the mirror. The **convex lens** shown in Figure 16.25 has been shaped so that all light rays that enter it parallel to its central axis cross one another at a single point on the opposite side of the lens. The central axis, or axis, is defined to be a line normal to the lens at its center. Such a lens is called a **converging lens** because of the converging effect it has on light rays. An expanded view of the path of one ray through the lens is shown in Figure 16.25 to illustrate how the ray changes

direction both as it enters and as it leaves the lens. Because the index of refraction of the lens is greater than that of air, the ray moves toward the perpendicular as it enters and away from the perpendicular as it leaves. (This is in accordance with the law of refraction.) As a result of the shape of the lens, light is thus bent toward the axis at both surfaces.

Figure 16.25 Rays of light entering a convex, or converging, lens parallel to its axis converge at its focal point, F. Ray 2 lies on the axis of the lens. The distance from the center of the lens to the focal point is the focal length, *f*, of the lens. An expanded view of the path taken by ray 1 shows the perpendiculars and the angles of incidence and refraction at both surfaces.

Note that rays from a light source placed at the focal point of a converging lens emerge parallel from the other side of the lens. You may have heard of the trick of using a converging lens to focus rays of sunlight to a point. Such a concentration of light energy can produce enough heat to ignite paper.

Figure 16.26 shows a **concave lens** and the effect it has on rays of light that enter it parallel to its axis (the path taken by ray 2 in the figure is the axis of the lens). The concave lens is a **diverging lens** because it causes the light rays to bend away (diverge) from its axis. In this case, the lens has been shaped so all light rays entering it parallel to its axis appear to originate from the same point, F, defined to be the focal point of a diverging lens. The distance from the center of the lens to the focal point is again called the *focal length*, or "*f*," of the lens. Note that the focal length of a diverging lens is defined to be negative. An expanded view of the path of one ray through the lens is shown in Figure 16.26 to illustrate how the shape of the lens, together with the law of refraction, causes the ray to follow its particular path and diverge.

Figure 16.26 Rays of light enter a concave, or diverging, lens parallel to its axis diverge and thus appear to originate from its focal point, F. The dashed lines are not rays; they indicate the directions from which the rays appear to come. The focal length, *f*, of a diverging lens is negative. An expanded view of the path taken by ray 1 shows the perpendiculars and the angles of incidence and refraction at both surfaces.

The power, *P*, of a lens is very easy to calculate. It is simply the reciprocal of the focal length, expressed in meters

$$P = \frac{1}{f}. \tag{16.15}$$

The units of power are diopters, D, which are expressed in reciprocal meters. If the focal length is negative, as it is for the diverging lens in Figure 16.26, then the power is also negative.

In some circumstances, a lens forms an image at an obvious location, such as when a movie projector casts an image onto a screen. In other cases, the image location is less obvious. Where, for example, is the image formed by eyeglasses? We use ray

tracing for thin lenses to illustrate how they form images, and we develop equations to describe the image-formation quantitatively. These are the rules for ray tracing:

1. A ray entering a converging lens parallel to its axis passes through the focal point, F, of the lens on the other side
2. A ray entering a diverging lens parallel to its axis seems to come from the focal point, F, on the side of the entering ray
3. A ray passing through the center of either a converging or a diverging lens does not change direction
4. A ray entering a converging lens through its focal point exits parallel to its axis
5. A ray that enters a diverging lens by heading toward the focal point on the opposite side exits parallel to the axis

Consider an object some distance away from a converging lens, as shown in Figure 16.27. To find the location and size of the image formed, we trace the paths of select light rays originating from one point on the object. In this example, the originating point is the top of a woman's head. Figure 16.27 shows three rays from the top of the object that can be traced using the ray-tracing rules just listed. Rays leave this point traveling in many directions, but we concentrate on only a few, which have paths that are easy to trace. The first ray is one that enters the lens parallel to its axis and passes through the focal point on the other side (rule 1). The second ray passes through the center of the lens without changing direction (rule 3). The third ray passes through the nearer focal point on its way into the lens and leaves the lens parallel to its axis (rule 4). All rays that come from the same point on the top of the person's head are refracted in such a way as to cross at the same point on the other side of the lens. The image of the top of the person's head is located at this point. Rays from another point on the object, such as the belt buckle, also cross at another common point, forming a complete image, as shown. Although three rays are traced in Figure 16.27, only two are necessary to locate the image. It is best to trace rays for which there are simple ray-tracing rules. Before applying ray tracing to other situations, let us consider the example shown in Figure 16.27 in more detail.

Figure 16.27 Ray tracing is used to locate the image formed by a lens. Rays originating from the same point on the object are traced. The three chosen rays each follow one of the rules for ray tracing, so their paths are easy to determine. The image is located at the point where the rays cross. In this case, a real image—one that can be projected on a screen—is formed.

The image formed in Figure 16.27 is a real image—meaning, it can be projected. That is, light rays from one point on the object actually cross at the location of the image and can be projected onto a screen, a piece of film, or the retina of an eye.

In Figure 16.27, the object distance, d_o, is greater than f. Now we consider a ray diagram for a convex lens where $d_o < f$, and another diagram for a concave lens.

Virtual Physics

Geometric Optics

Click to view content (https://www.openstax.org/l/28Geometric)

This animation shows you how the image formed by a convex lens changes as you change object distance, curvature radius, refractive index, and diameter of the lens. To begin, choose Principal Rays in the upper left menu and then try varying some of the parameters indicated at the upper center. Show Help supplies a few helpful labels.

How does the focal length, f, change with an increasing radius of curvature? How does f change with an increasing refractive index?

a. The focal length increases in both cases: when the radius of curvature and the refractive index increase.
b. The focal length decreases in both cases: when the radius of curvature and the refractive index increase.
c. The focal length increases when the radius of curvature increases; it decreases when the refractive index increases.
d. The focal length decreases when the radius of curvature increases; it increases in when the refractive index increases.

Type	Formed When	Image Type	d_i	M
Case 1	f positive, $d_o > f$	Real	Positive	Negative $m >, <,$ or $= -1$
Case 2	f positive, $d_o < f$	Virtual	Negative	Positive $m > 1$
Case 3	f negative	Virtual	Negative	Positive $m < 1$

Table 16.3 Three Types of Images Formed by Lenses

The examples in Figure 16.27 and Figure 16.28 represent the three possible cases—case 1, case 2, and case 3—summarized in Table 16.3. In the table, m is magnification; the other symbols have the same meaning as they did for curved mirrors.

Figure 16.28 (a) The image is virtual and larger than the object. (b) The image is virtual and smaller than the object.

Snap Lab

Focal Length

- Temperature extremes—Very hot or very cold temperatures are encountered in this lab that can cause burns. Use protective mitts, eyewear, and clothing when handling very hot or very cold objects. Notify your teacher immediately of any burns.
- EYE SAFETY—Looking at the Sun directly can cause permanent eye damage. Do not look at the Sun through any lens.

- Several lenses
- A sheet of white paper
- A ruler or tape measure

Instructions

Procedure

1. Find several lenses and determine whether they are converging or diverging. In general, those that are thicker near the edges are diverging and those that are thicker near the center are converging.
2. On a bright, sunny day take the converging lenses outside and try focusing the sunlight onto a sheet of white paper.
3. Determine the focal lengths of the lenses. Have one partner slowly move the lens toward and away from the paper until you find the distance at which the light spot is at its brightest. Have the other partner measure the distance from the lens to the bright spot. Be careful, because the paper may start to burn, depending on the type of lens.

True or false—The bright spot that appears in focus on the paper is an image of the Sun.

a. True
b. False

Image formation by lenses can also be calculated from simple equations. We learn how these calculations are carried out near the end of this section.

Some common applications of lenses with which we are all familiar are magnifying glasses, eyeglasses, cameras, microscopes, and telescopes. We take a look at the latter two examples, which are the most complex. We have already seen the design of a telescope that uses only mirrors in . Figure 16.29 shows the design of a telescope that uses two lenses. Part (a) of the figure shows the design of the telescope used by Galileo. It produces an upright image, which is more convenient for many applications. Part (b) shows an arrangement of lenses used in many astronomical telescopes. This design produces an inverted image, which is less of a problem when viewing celestial objects.

Figure 16.29 (a) Galileo made telescopes with a convex objective and a concave eyepiece. They produce an upright image and are used in spyglasses. (b) Most simple telescopes have two convex lenses. The objective forms a case 1 image, which is the object for the eyepiece. The eyepiece forms a case 2 final image that is magnified.

Figure 16.30 shows the path of light through a typical microscope. Microscopes were first developed during the early 1600s by eyeglass makers in the Netherlands and Denmark. The simplest compound microscope is constructed from two convex lenses, as shown schematically in Figure 16.30. The first lens is called the *objective lens*; it has typical magnification values from 5 × to 100 × . In standard microscopes, the **objectives** are mounted such that when you switch between them, the sample remains in focus. Objectives arranged in this way are described as ***parfocal***. The second lens, the **eyepiece**, also referred to as the ***ocular***, has several lenses that slide inside a cylindrical barrel. The focusing ability is provided by the movement of both the objective lens and the eyepiece. The purpose of a microscope is to magnify small objects, and both lenses contribute to the final magnification. In addition, the final enlarged image is produced in a location far enough from the observer to be viewed easily because the eye cannot focus on objects or images that are too close.

Figure 16.30 A compound microscope composed of two lenses, an objective and an eyepiece. The objective forms a case 1 image that is larger than the object. This first image is the object for the eyepiece. The eyepiece forms a case 2 final image that is magnified even further.

Real lenses behave somewhat differently from how they are modeled using rays diagrams or the thin-lens equations. Real lenses produce aberrations. An **aberration** is a distortion in an image. There are a variety of aberrations that result from lens size, material, thickness, and the position of the object. One common type of aberration is **chromatic aberration**, which is related to color. Because the index of refraction of lenses depends on color, or wavelength, images are produced at different places and with different magnifications for different colors. The law of reflection is independent of wavelength, so mirrors do not have this problem. This result is another advantage for the use of mirrors in optical systems such as telescopes.

Figure 16.31(a) shows chromatic aberration for a single convex lens, and its partial correction with a two-lens system. The index of refraction of the lens increases with decreasing wavelength, so violet rays are refracted more than red rays, and are thus focused closer to the lens. The diverging lens corrects this in part, although it is usually not possible to do so completely. Lenses made of different materials and with different dispersions may be used. For example, an achromatic doublet consisting of a converging lens made of crown glass in contact with a diverging lens made of flint glass can reduce chromatic aberration dramatically (Figure 16.31(b)).

Figure 16.31 (a) Chromatic aberration is caused by the dependence of a lens's index of refraction on color (wavelength). The lens is more powerful for violet (V) than for red (R), producing images with different colors, locations, and magnifications. (b) Multiple-lens systems can correct chromatic aberrations in part, but they may require lenses of different materials and add to the expense of optical systems such as cameras.

Physics of the Eye

The eye is perhaps the most interesting of all optical instruments. It is remarkable in how it forms images and in the richness of detail and color they eye can detect. However, our eyes commonly need some correction to reach what is called *normal* vision, but should be called *ideal* vision instead. Image formation by our eyes and common vision correction are easy to analyze using geometric optics. Figure 16.32 shows the basic anatomy of the eye. The cornea and lens form a system that, to a good approximation, acts as a single thin lens. For clear vision, a real image must be projected onto the light-sensitive retina, which lies at a fixed distance from the lens. The lens of the eye adjusts its power to produce an image on the retina for objects at different distances. The center of the image falls on the fovea, which has the greatest density of light receptors and the greatest acuity (sharpness) in the visual field. There are no receptors at the place where the optic nerve meets the eye, which is called the *blind spot*. An image falling on this spot cannot be seen. The variable opening (or pupil) of the eye along with chemical adaptation allows the eye to detect light intensities from the lowest observable to 10^{10} times greater (without damage). Ten orders of magnitude is an incredible range of detection. Our eyes perform a vast number of functions, such as sense direction, movement, sophisticated colors, and distance. Processing of visual nerve impulses begins with interconnections in the retina and continues in the brain. The optic nerve conveys signals received by the eye to the brain.

Figure 16.32 The cornea and lens of an eye act together to form a real image on the light-sensing retina, which has its densest concentration of receptors in the fovea, and a blind spot over the optic nerve. The power of the lens of an eye is adjustable to provide an image on the retina for varying object distances.

Refractive indices are crucial to image formation using lenses. Table 16.4 shows refractive indices relevant to the eye. The biggest change in the refractive index—and the one that causes the greatest bending of rays—occurs at the cornea rather than the lens. The ray diagram in Figure 16.33 shows image formation by the cornea and lens of the eye. The rays bend according to the refractive indices provided in Table 16.4. The cornea provides about two-thirds of the magnification of the eye because the speed of light changes considerably while traveling from air into the cornea. The lens provides the remaining magnification needed to produce an image on the retina. The cornea and lens can be treated as a single thin lens, although the light rays pass through several layers of material (such as the cornea, aqueous humor, several layers in the lens, and vitreous humor), changing direction at each interface. The image formed is much like the one produced by a single convex lens. This result is a case 1 image. Images formed in the eye are inverted, but the brain inverts them once more to make them seem upright.

Material	Index of Refraction
Water	1.33
Air	1.00
Cornea	1.38
Aqueous humor	1.34

*The index of refraction varies throughout the lens and is greatest at its center.

Table 16.4 Refractive Indices Relevant to the Eye

Material	Index of Refraction
Lens	1.41 average*
Vitreous humor	1.34

*The index of refraction varies throughout the lens and is greatest at its center.

Table 16.4 Refractive Indices Relevant to the Eye

Figure 16.33 An image is formed on the retina, with light rays converging most at the cornea and on entering and exiting the lens. Rays from the top and bottom of the object are traced and produce an inverted real image on the retina. The distance to the object is drawn smaller than scale.

As noted, the image must fall precisely on the retina to produce clear vision—that is, the image distance, d_i, must equal the lens-to-retina distance. Because the lens-to-retina distance does not change, d_i must be the same for objects at all distances. The eye manages to vary the distance by varying the power (and focal length) of the lens to accommodate for objects at various distances. In Figure 16.33, you can see the small ciliary muscles above and below the lens that change the shape of the lens and, thus, the focal length.

The need for some type of vision correction is very common. Common vision defects are easy to understand, and some are simple to correct. Figure 16.34 illustrates two common vision defects. Nearsightedness, or myopia, is the inability to see distant objects clearly while close objects are in focus. The nearsighted eye *overconverges* the nearly parallel rays from a distant object, and the rays cross in front of the retina. More divergent rays from a close object are converged on the retina, producing a clear image. Farsightedness, or hyperopia, is the inability to see close objects clearly whereas distant objects may be in focus. A farsighted eye does not converge rays from a close object sufficiently to make the rays meet on the retina. Less divergent rays from a distant object can be converged for a clear image.

Figure 16.34 (a) The nearsighted (myopic) eye converges rays from a distant object in front of the retina; thus, they are diverging when they strike the retina, and produce a blurry image. This divergence can be caused by the lens of the eye being too powerful (in other words, too short a focal length) or the length of the eye being too great. (b) The farsighted (hyperopic) eye is unable to converge the rays from a close object by the time they strike the retina and produce ... blurry close vision. This poor convergence can be caused by insufficient power (in

other words, too long a focal length) in the lens or by the eye being too short.

Because the nearsighted eye overconverges light rays, the correction for nearsightedness involves placing a diverging spectacle lens in front of the eye. This lens reduces the power of an eye that has too short a focal length (Figure 16.35(a)). Because the farsighted eye *underconverges* light rays, the correction for farsightedness is to place a converging spectacle lens in front of the eye. This lens increases the power of an eye that has too long a focal length (Figure 16.35(b)).

Figure 16.35 (a) Correction of nearsightedness requires a diverging lens that compensates for the overconvergence by the eye. The diverging lens produces an image closer to the eye than the object so that the nearsighted person can see it clearly. (b) Correction of farsightedness uses a converging lens that compensates for the underconvergence by the eye. The converging lens produces an image farther from the eye than the object so that the farsighted person can see it clearly. In both (a) and (b), the rays that meet at the retina represent corrected vision, and the other rays represent blurred vision without corrective lenses.

Calculations Using Lens Equations

As promised, there are no new equations to memorize. We can use equations already presented for solving problems involving curved mirrors. Careful analysis allows you to apply these equations to lenses. Here are the equations you need

$$P = \frac{1}{f},$$

where P is power, expressed in reciprocal meters (m^{-1}) rather than diopters (D), and f is focal length, expressed in meters (m). You also need

$$\frac{1}{f} = \frac{1}{d_i} + \frac{1}{d_o},$$

where, as before, d_o and d_i are object distance and image distance, respectively. Remember, this equation is usually more useful if rearranged to solve for one of the variables. For example,

$$d_i = \frac{f d_o}{d_o - f}.$$

The equations for magnification, m, are also the same as for mirrors

$$m = \frac{h_i}{h_o} = -\frac{d_i}{d_o},$$

where h_i and h_o are the image height and object height, respectively. Remember, also, that a negative d_i value indicates a virtual image and a negative h_i value indicates an inverted image.

These are the steps to follow when solving a lens problem:

- Step 1. Examine the situation to determine that image formation by a lens is involved.
- Step 2. Determine whether ray tracing, the thin-lens equations, or both should be used. A sketch is very helpful even if ray tracing is not specifically required by the problem. Write useful symbols and values on the sketch.
- Step 3. Identify exactly what needs to be determined in the problem (identify the unknowns).
- Step 4. Make a list of what is given or can be inferred from the problem as stated (identify the knowns). It is helpful to determine whether the situation involves a case 1, 2, or 3 image. Although these are just names for types of images, they have certain characteristics (given in Table 16.3) that can be of great use in solving problems.

- Step 5. If ray tracing is required, use the ray-tracing rules listed earlier in this section.
- Step 6. Most quantitative problems require the use of the thin-lens equations. These equations are solved in the usual manner by substituting knowns and solving for unknowns. Several worked examples were included earlier and can serve as guides.
- Step 7. Check whether the answer is reasonable. Does it make sense? If you identified the type of image (case 1, 2, or 3) correctly, you should assess whether your answer is consistent with the type of image, magnification, and so on.

All problems will be solved by one or more of the equations just presented, with ray tracing used only for general analysis of the problem. The steps then simplify to the following:

1. Identify the unknown.
2. Identify the knowns.
3. Choose an equation, plug in the knowns, and solve for the unknown.

Here are some worked examples:

WORKED EXAMPLE

The Power of a Magnifying Glass

Strategy

The Sun is so far away that its rays are nearly parallel when they reach Earth. The magnifying glass is a convex (or converging) lens, focusing the nearly parallel rays of sunlight. Thus, the focal length of the lens is the distance from the lens to the spot, and its power, in diopters (D), is the inverse of this distance (in reciprocal meters).

Solution

The focal length of the lens is the distance from the center of the lens to the spot, which we know to be 8.00 cm. Thus,

$$f = 8.00 \text{ cm}. \tag{16.16}$$

To find the power of the lens, we must first convert the focal length to meters; then, we substitute this value into the equation for power.

$$P = \frac{1}{f} = \frac{1}{0.0800 \text{ m}} = 12.5 \text{ D} \tag{16.17}$$

Discussion

This result demonstrates a relatively powerful lens. Remember that the power of a lens in diopters should not be confused with the familiar concept of power in watts.

WORKED EXAMPLE

Image Formation by a Convex Lens

A clear glass light bulb is placed 0.75 m from a convex lens with a 0.50 m focal length, as shown in Figure 16.36. Use ray tracing to get an approximate location for the image. Then, use the mirror/lens equations to calculate (a) the location of the image and (b) its magnification. Verify that ray tracing and the thin-lens and magnification equations produce consistent results.

Figure 16.36 A light bulb placed 0.75 m from a lens with a 0.50 m focal length produces a real image on a poster board, as discussed in the previous example. Ray tracing predicts the image location and size.

Strategy

Because the object is placed farther away from a converging lens than the focal length of the lens, this situation is analogous to the one illustrated in the previous figure of a series of drawings showing a woman standing to the left of a lens. Ray tracing to scale should produce similar results for d_i. Numerical solutions for d_i and m can be obtained using the thin-lens and magnification equations, noting that $d_o = 0.75$ m and $f = 0.50$ m.

Solution

The ray tracing to scale in Figure 16.36 shows two rays from a point on the bulb's filament crossing about 1.50 m on the far side of the lens. Thus, the image distance, d_i, is about 1.50 m. Similarly, the image height based on ray tracing is greater than the object height by about a factor of two, and the image is inverted. Thus, m is about −2. The minus sign indicates the image is inverted. The lens equation can be rearranged to solve for d_i from the given information.

$$d_i = \frac{fd_o}{d_o - f} = \frac{(0.50)(0.75)}{0.75–0.50} = 1.5 \text{ m} \tag{16.18}$$

Now, we use $\frac{d_i}{d_o}$ to find m.

$$m = -\frac{d_i}{d_o} = -\frac{1.5}{0.75} = -2.0 \tag{16.19}$$

Discussion

Note that the minus sign causes the magnification to be negative when the image is inverted. Ray tracing and the use of the lens equation produce consistent results. The thin-lens equation gives the most precise results, and is limited only by the accuracy of the given information. Ray tracing is limited by the accuracy with which you draw, but it is highly useful both conceptually and visually.

WORKED EXAMPLE

Image Formation by a Concave Lens

Suppose an object, such as a book page, is held 6.50 cm from a concave lens with a focal length of −10.0 cm. Such a lens could be used in eyeglasses to correct pronounced nearsightedness. What magnification is produced?

Strategy

This example is identical to the preceding one, except that the focal length is negative for a concave or diverging lens. The method of solution is therefore the same, but the results are different in important ways.

Solution

$$d_i = \frac{fd_o}{d_o - f} = \frac{(-10.0)(6.50)}{6.50 - (-10.0)} = -3.94 \text{ cm} \tag{16.20}$$

Now the magnification equation can be used to find the magnification, m, because both d_i and d_o are known. Entering their values gives

$$m = -\frac{d_i}{d_o} = -\frac{-3.94}{6.50} = 0.606. \tag{16.21}$$

Discussion

A number of results in this example are true of all case 3 images. Magnification is positive (as calculated), meaning the image is upright. The magnification is also less than one, meaning the image is smaller than the object—in this case, a little more than half its size. The image distance is negative, meaning the image is on the same side of the lens as the object. The image is virtual. The image is closer to the lens than the object, because the image distance is smaller in magnitude than the object distance. The location of the image is not obvious when you look through a concave lens. In fact, because the image is smaller than the object, you may think it is farther away; however, the image is closer than the object—a fact that is useful in correcting nearsightedness.

WATCH PHYSICS

The Lens Equation and Problem Solving

The video shows calculations for both concave and convex lenses. It also explains real versus virtual images, erect versus inverted images, and the significance of negative and positive signs for the involved variables.

Click to view content (https://www.openstax.org/l/28Lenses)

If a lens has a magnification of $-\frac{1}{2}$, how does the image compare with the object in height and orientation?

a. The image is erect and is half as tall as the object.
b. The image is erect and twice as tall as the object.
c. The image is inverted and is half as tall as the object.
d. The image is inverted and is twice as tall as the object.

Practice Problems

11. A lens has a focal length of 12.5 cm. What is the power of the lens?

a. The power of the lens is 0.0400 D.
b. The power of the lens is 0.0800 D.
c. The power of the lens is 4.00 D.
d. The power of the lens is 8.00 D.

12. If a lens produces a 5.00 -cm tall image of an 8.00 -cm -high object when placed 10.0 cm from the lens, what is the apparent image distance? Construct a ray diagram using paper, a pencil, and a ruler to confirm your calculation.

a. −3.12 cm
b. −6.25 cm
c. 3.12 cm
d. 6.25 cm

Check Your Understanding

13. A lens has a magnification that is negative. What is the orientation of the image?

a. Negative magnification means the image is erect and real.
b. Negative magnification means the image is erect and virtual.
c. Negative magnification means the image is inverted and virtual.
d. Negative magnification means the image is inverted and real.

14. Which part of the eye controls the amount of light that enters?

a. the pupil
b. the iris
c. the cornea
d. the retina

15. An object is placed between the focal point and a convex lens. Describe the image that is formed in terms of its orientation, and whether the image is real or virtual.

a. The image is real and erect.
b. The image is real and inverted.
c. The image is virtual and erect.
d. The image is virtual and inverted.

16. A farsighted person buys a pair of glasses to correct her farsightedness. Describe the main symptom of farsightedness and the type of lens that corrects it.

a. Farsighted people cannot focus on objects that are far away, but they can see nearby objects easily. A convex lens is used to correct this.
b. Farsighted people cannot focus on objects that are close up, but they can see far-off objects easily. A concave lens is used to correct this.

c. Farsighted people cannot focus on objects that are close up, but they can see distant objects easily. A convex lens is used to correct this.
d. Farsighted people cannot focus on objects that are either close up or far away. A concave lens is used to correct this.

KEY TERMS

aberration a distortion in an image produced by a lens
angle of incidence the angle, with respect to the normal, at which a ray meets a boundary between media or a reflective surface
angle of reflection the angle, with respect to the normal, at which a ray leaves a reflective surface
angle of refraction the angle between the normal and the refracted ray
central axis a line perpendicular to the center of a lens or mirror extending in both directions
chromatic aberration an aberration related to color
concave lens a lens that causes light rays to diverge from the central axis
concave mirror a mirror with a reflective side that is curved inward
converging lens a convex lens
convex lens a lens that causes light rays to converge toward the central axis
convex mirror a mirror with a reflective side that is curved outward
critical angle an incident angle that produces an angle of refraction of 90°
dispersion separation of white light into its component wavelengths
diverging lens a concave lens
focal length the distance from the focal point to the mirror
focal point the point at which rays converge or appear to converge
incident ray the incoming ray toward a medium boundary or a reflective surface
index of refraction the speed of light in a vacuum divided by the speed of light in a given material
law of reflection the law that indicates the angle of reflection equals the angle of incidence
law of refraction the law that describes the relationship between refractive indices of materials on both sides of a boundary and the change in the path of light crossing the boundary, as given by the equation $n_1 \sin \theta_1 = n_2 \sin \theta_2$
ray light traveling in a straight line
real image an optical image formed when light rays converge and pass through the image, producing an image that can be projected onto a screen
refracted ray the light ray after it has been refracted
Snell's law the law of refraction expressed mathematically as $n_1 \sin \theta_1 = n_2 \sin \theta_2$
total internal reflection reflection of light traveling through a medium with a large refractive index at a boundary of a medium with a low refractive index under conditions such that refraction cannot occur
virtual image the point from which light rays appear to diverge without actually doing so

SECTION SUMMARY

16.1 Reflection

- The angle of reflection equals the angle of incidence.
- Plane mirrors and convex mirrors reflect virtual, erect images. Concave mirrors reflect light to form real, inverted images or virtual, erect images, depending on the location of the object.
- Image distance, height, and other characteristics can be calculated using the lens/mirror equation and the magnification equation.

16.2 Refraction

- The index of refraction for a material is given by the speed of light in a vacuum divided by the speed of light in that material.
- Snell's law states the relationship between indices of refraction, the incident angle, and the angle of refraction.
- The critical angle, θ_c, determines whether total internal refraction can take place, and can be calculated according to $\theta_c = sin^{-1}(\frac{n_2}{n_1})$.

16.3 Lenses

- The characteristics of images formed by concave and convex lenses can be predicted using ray tracing. Characteristics include real versus virtual, inverted versus upright, and size.
- The human eye and corrective lenses can be explained using geometric optics.
- Characteristics of images formed by lenses can be calculated using the mirror/lens equation.

KEY EQUATIONS

16.1 Reflection

lens/mirror equation (reciprocal version) $\frac{1}{f} = \frac{1}{d_i} + \frac{1}{d_o}$

lens/mirror equation (solved version) $f = \frac{d_i d_o}{d_o + d_i}$

magnification equation	$m = \frac{h_i}{h_o} = \frac{-d_i}{d_o}$
radius/focal length equation	$R = 2f$

16.2 Refraction

index of Refraction	$n = \frac{c}{v}$
Snell's law	$n_1 \sin\theta_1 = n_2 \sin\theta_2$ or $\frac{n_1}{n_2} = \frac{\sin\theta_2}{\sin\theta_1}$
>Snell's law in terms of speed	$\frac{\sin\theta_1}{\sin\theta_2} = \frac{v_1}{v_2}$
critical angle	$\theta_c = \sin^{-1}(\frac{n_2}{n_1})$, for $n_1 > n_2$

16.3 Lenses

power and focal length	$P = \frac{1}{f}$
mirror/lens (or thin-lens) equation	$\frac{1}{f} = \frac{1}{d_i} + \frac{1}{d_o}$
rearranged mirror/lens equation	$d_i = \frac{fd_o}{d_o - f}$
magnification equation	$m = \frac{h_i}{h_o} = -\frac{d_i}{d_o}$

CHAPTER REVIEW

Concept Items

16.1 Reflection

1. Part A. Can you see a virtual image? Part B. Can you photograph one? Explain your answers.
   a. A. yes; B. No, an image from a flat mirror cannot be photographed.
   b. A. no; B. Yes, an image from a flat mirror can be photographed.
   c. A. yes; B. Yes, an image from a flat mirror can be photographed.
   d. A. no; B. No, an image from a flat mirror cannot be photographed.

2. State the law of reflection.
   a. $\vartheta_r = \vartheta_i$, where ϑ_r is the angle of reflection and ϑ_i is the angle of incidence.
   b. $\vartheta_r > \vartheta_i$, where ϑ_r is the angle of reflection and ϑ_i is the angle of incidence.
   c. $\vartheta_r < \vartheta_i$, where ϑ_r is the angle of reflection and ϑ_i is the angle of incidence.
   d. $\vartheta_r = 0$, where ϑ_r is the angle of reflection.

16.2 Refraction

3. Does light change direction toward or away from the normal when it goes from air to water? Explain.
   a. The light bends away from the normal because the index of refraction of water is greater than that of air.
   b. The light bends away from the normal because the index of refraction of air is greater than that of water.
   c. The light bends toward the normal because the index of refraction of water is greater than that of air.
   d. The light bends toward the normal because the index of refraction of air is greater than that of water.

16.3 Lenses

4. An object is positioned in front of a lens with its base resting on the principal axis. Describe two rays that could be traced from the top of the object and through the lens that would locate the top of an image.
   a. A ray perpendicular to the axis and a ray through the center of the lens
   b. A ray parallel to the axis and a ray that does not pass through the center of the lens
   c. A ray parallel to the axis and a ray through the center of the lens
   d. A ray parallel to the axis and a ray that does not pass through the focal point

5. A person timing the moonrise looks at her watch and then at the rising moon. Describe what happened inside her eyes that allowed her to see her watch clearly one second and then see the moon clearly.
   a. The shape of the lens was changed by the sclera, and thus its focal length was also changed, so that each of the images focused on the retina.
   b. The shape of the lens was changed by the choroid, and thus its focal length was also changed, so that each of the images focused on the retina.
   c. The shape of the lens was changed by the iris, and thus its focal length was also changed, so that each of the images focused on the retina.

d. The shape of the lens was changed by the muscles, and thus its focal length was also changed, so that each of the images focused on the retina.

6. For a concave lens, if the image distance, d_i, is negative, where does the image appear to be with respect to the object?
   a. The image always appears on the same side of the lens.
   b. The image appears on the opposite side of the lens.
   c. The image appears on the opposite side of the lens only if the object distance is greater than the focal length.
   d. The image appears on the same side of the lens only if the object distance is less than the focal length.

Critical Thinking Items

16.1 Reflection

7. Why are diverging mirrors often used for rear-view mirrors in vehicles? What is the main disadvantage of using such a mirror compared with a flat one?
   a. It gives a wide range of view. The image appears to be closer than the actual object.
   b. It gives a narrow range of view. The image appears to be farther than the actual object.
   c. It gives a narrow range of view. The image appears to be closer than the actual object.
   d. It gives a wide range of view. The image appears to be farther than the actual object.

16.2 Refraction

8. A high-quality diamond may be quite clear and colorless, transmitting all visible wavelengths with little absorption. Explain how it can sparkle with flashes of brilliant color when illuminated by white light.
   a. Diamond and air have a small difference in their refractive indices that results in a very small critical angle. The light that enters a diamond may exit at only a few points, and these points sparkle because many rays have been directed toward them.
   b. Diamond and air have a small difference in their refractive indices that results in a very large critical angle. The light that enters a diamond may exit at only a few points, and these points sparkle because many rays have been directed toward them.
   c. Diamond has a high index of refraction with respect to air, which results in a very small critical angle. The light that enters a diamond may exit at only a few points, and these points sparkle because many rays have been directed toward them.
   d. Diamond has a high index of refraction with respect to air, which results in a very large critical angle. The light that enters a diamond may exit at only a few points, and these points sparkle because many rays have been directed toward them.

9. The most common type of mirage is an illusion in which light from far-away objects is reflected by a pool of water that is not really there. Mirages are generally observed in deserts, where there is a hot layer of air near the ground. Given that the refractive index of air is less for air at higher temperatures, explain how mirages can be formed.
   a. The hot layer of air near the ground is lighter than the cooler air above it, but the difference in refractive index is small, which results in a large critical angle. The light rays coming from the horizon strike the hot air at large angles, so they are reflected as they would be from water.
   b. The hot layer of air near the ground is lighter than the cooler air above it, and the difference in refractive index is large, which results in a large critical angle. The light rays coming from the horizon strike the hot air at large angles, so they are reflected as they would be from water.
   c. The hot layer of air near the ground is lighter than the cooler air above it, but the difference in refractive index is small, which results in a small critical angle. The light rays coming from the horizon strike the hot air at large angles, so they are reflected as they would be from water.
   d. The hot layer of air near the ground is lighter than the cooler air above it, and the difference in the refractive index is large, which results in a small critical angle. The light rays coming from the horizontal strike the hot air at large angles, so they are reflected as they would be from water.

16.3 Lenses

10. When you focus a camera, you adjust the distance of the lens from the film. If the camera lens acts like a thin lens, why can it not be kept at a fixed distance from the film for both near and distant objects?
    a. To focus on a distant object, you need to increase the image distance.
    b. To focus on a distant object, you need to increase the focal length of the lens.
    c. To focus on a distant object, you need to decrease the focal length of the lens.
    d. To focus on a distant object, you may need to increase or decrease the focal length of the lens.

11. Part A—How do the refractive indices of the cornea,

aqueous humor, and the lens of the eye compare with the refractive index of air?
Part B—How do the comparisons in part A explain how images are focused on the retina?

a. (A) The cornea, aqueous humor, and lens of the eye have smaller refractive indices than air.
(B) Rays entering the eye are refracted away from the central axis, which causes them to meet at the focal point on the retina.
b. (A) The cornea, aqueous humor, and lens of the eye have greater refractive indices than air.
(B) Rays entering the eye are refracted away from the central axis, which causes them to meet at the focal point on the retina.
c. (A) The cornea, aqueous humor, and lens of the eye have smaller refractive indices than air.
(B) Rays entering the eye are refracted toward the central axis, which causes them to meet at the focal point on the retina.
d. (A) The cornea, aqueous humor, and lens of the eye have greater refractive indices than air.
(B) Rays entering the eye are refracted toward the central axis, which causes them to meet at the focal point on the retina.

Problems

16.1 Reflection

12. Some telephoto cameras use a mirror rather than a lens. What radius of curvature is needed for a concave mirror to replace a 0.800 -m focal-length telephoto lens?

a. 0.400 m
b. 1.60 m
c. 4.00 m
d. 16.0 m

13. What is the focal length of a makeup mirror that produces a magnification of 2.00 when a person's face is 8.00 cm away?

a. −16 cm
b. −5.3 cm
c. 5.3 cm
d. 16 cm

16.2 Refraction

14. An optical fiber uses flint glass ($n = 1.66$) clad with crown glass ($n = 1.52$) . What is the critical angle?

a. 33.2°
b. 23.7°
c. 0.92 rad
d. 1.16 rad

15. Suppose this figure represents a ray of light going from air ($n = 1.0003$) through crown glass ($n = 1.52$) into water, similar to a beam of light going into a fish tank.

Calculate the amount the ray is displaced by the glass (Δx), given that the incident angle is 40.0° and the glass is 1.00 cm thick.

a. 0.839 cm
b. 0.619 cm
c. 0.466 cm
d. 0.373 cm

16.3 Lenses

16. A camera's zoom lens has an adjustable focal length ranging from 80.0 to 200 mm . What is its range of powers?

a. The lowest power is 0.05 D and the highest power is 0.125 D.
b. The lowest power is 0.08 D and the highest power is 0.20 D.
c. The lowest power is 5.00 D and the highest power is 12.5 D.
d. The lowest power is 80 D and the highest power is 200 D.

17. Suppose a telephoto lens with a focal length of 200 mm is being used to photograph mountains 10.0 km away. (a) Where is the image? (b) What is the height of the image of a 1,000-m-high cliff on one of the mountains?

a. (a) The image is 0.200 m on the same side of the lens. (b) The height of the image is – 2.00 cm.
b. (a) The image is 0.200 m on the opposite side of the

lens. (b) The height of the image is – 2.00 cm.

c. (a) The image is 0.200 m on the opposite side of the lens. (b) The height of the image is +2.00 cm.

d. (a) The image is 0.100 m on the same side of the lens. (b) The height of the image is +2.00 cm.

Performance Task

16.3 Lenses

18. In this performance task, you will investigate the lens-like properties of a clear bottle.

- a water bottle or glass with a round cross-section and smooth, vertical sides
- enough water to fill the bottle
- a meter stick or tape measure
- a bright light source with a small bulb, such as a pen light
- a small bright object, such as a silver spoon.

Instructions

Procedure

1. Look through a clear glass or plastic bottle and describe what you see.
2. Next, fill the bottle with water and describe what you see.
3. Use the water bottle as a lens to produce the image of a bright object.
4. Estimate the focal length of the water bottle lens.
   a. How can you find the focal length of the lens using the light and a blank wall?
   b. How can you find the focal length of the lens using the bright object?
   c. Why did the water change the lens properties of the bottle?

TEST PREP

Multiple Choice

16.1 Reflection

19. In geometric optics, a straight line emerging from a point is called a (an) _____.

a. ray
b. focal point
c. image
d. object distance

20. An image of a 2.0 -cm object reflected from a mirror is 5.0 cm tall. What is the magnification of the mirror?

a. 0.4
b. 2.5
c. 3
d. 10

21. Can a virtual image be projected onto a screen with additional lenses or mirrors? Explain your answer.

a. Yes, the rays actually meet behind the lens or mirror.
b. No, the image is formed by rays that converge to a point in front of the mirror or lens.
c. Yes, any image that can be seen can be manipulated so that it can be projected onto a screen.
d. No, the image can only be perceived as being behind the lens or mirror.

16.2 Refraction

22. What does c represent in the equation $n = \frac{c}{v}$?

a. the critical angle
b. the refractive index
c. the speed of light in a vacuum
d. the speed of light in a transparent material

23. What is the term for the minimum angle at which a light ray is reflected back into a material and cannot pass into the surrounding medium?

a. critical angle
b. incident angle
c. angle of refraction
d. angle of reflection

24. Consider these indices of refraction: glass: 1.52, air: 1.0003, water: 1.333. Put these materials in order from the one in which the speed of light is fastest to the one in which it is slowest.

a. The speed of light in water > the speed of light in air > the speed of light in glass.
b. The speed of light in glass > the speed of light in water > the speed of light in air.
c. The speed of light in air > the speed of light in water > the speed of light in glass.
d. The speed of light in glass > the speed of light in air > the speed of light in water.

25. Explain why an object in water always appears to be at a depth that is more shallow than it actually is.

a. Because of the refraction of light, the light coming from the object bends toward the normal at the interface of water and air. This causes the object to appear at a location that is above the actual position of the object. Hence, the image appears to

be at a depth that is more shallow than the actual depth.

b. Because of the refraction of light, the light coming from the object bends away from the normal at the interface of water and air. This causes the object to appear at a location that is above the actual position of the object. Hence, the image appears to be at a depth that is more shallow than the actual depth.

c. Because of the refraction of light, the light coming from the object bends toward the normal at the interface of water and air. This causes the object to appear at a location that is below the actual position of the object. Hence, the image appears to be at a depth that is more shallow than the actual depth.

d. Because of the refraction of light, the light coming from the object bends away from the normal at the interface of water and air. This causes the object to appear at a location that is below the actual position of the object. Hence, the image appears to be at a depth that is more shallow than the actual depth.

16.3 Lenses

26. For a given lens, what is the height of the image divided by the height of the object ($\frac{h_i}{h_o}$) equal to?

a. power
b. focal length
c. magnification
d. radius of curvature

27. Which part of the eye has the greatest density of light receptors?

a. the lens
b. the fovea
c. the optic nerve
d. the vitreous humor

28. What is the power of a lens with a focal length of 10 cm?

a. 10 m^{-1}, or 10 D
b. 10 cm^{-1}, or 10 D
c. 10 m, or 10 D
d. 10 cm, or 10 D

29. Describe the cause of chromatic aberration.

a. Chromatic aberration results from the dependence of the frequency of light on the refractive index, which causes dispersion of different colors of light by a lens so that each color has a different focal point.

b. Chromatic aberration results from the dispersion of different wavelengths of light by a curved mirror so that each color has a different focal point.

c. Chromatic aberration results from the dependence of the reflection angle at a spherical mirror's surface on the distance of light rays from the principal axis so that different colors have different focal points.

d. Chromatic aberration results from the dependence of the wavelength of light on the refractive index, which causes dispersion of different colors of light by a lens so that each color has a different focal point.

Short Answer

16.1 Reflection

30. Distinguish between reflection and refraction in terms of how a light ray changes when it meets the interface between two media.

a. Reflected light penetrates the surface whereas refracted light is *bent* as it travels from one medium to the other.

b. Reflected light penetrates the surface whereas refracted light travels along a curved path.

c. Reflected light *bounces* from the surface whereas refracted light travels along a curved path.

d. Reflected light *bounces* from the surface whereas refracted light is *bent* as it travels from one medium to the other.

31. Sometimes light may be both reflected and refracted as it meets the surface of a different medium. Identify a material with a surface that when light travels through the air it is both reflected and refracted. Explain how this is possible.

a. Light passing through air is partially reflected and refracted when it meets a glass surface. It is reflected because glass has a smooth surface; it is refracted while passing into the transparent glass.

b. Light passing through air is partially reflected and refracted when it meets a glass surface. It is reflected because glass has a rough surface, and it is refracted while passing into the opaque glass.

c. Light passing through air is partially reflected and refracted when it meets a glass surface. It is reflected because glass has a smooth surface; it is refracted while passing into the opaque glass.

d. Light passing through air is partially reflected and refracted when it meets a glass surface. It is reflected because glass has a rough surface; it is refracted while passing into the transparent glass.

32. A concave mirror has a focal length of 5.00 cm. What is

the image distance of an object placed 7.00 cm from the center of the mirror?

a. −17.5 cm
b. −2.92 cm
c. 2.92 cm
d. 17.5 cm

33. An 8.0 -cm tall object is placed 6.0 cm from a concave mirror with a magnification of −2.0. What are the image height and the image distance?

a. $h_i = -16$ cm, $d_i = -12$ cm
b. $h_i = -16$ cm, $d_i = 12$ cm
c. $h_i = 16$ cm, $d_i = -12$ cm
d. $h_i = 16$ cm, $d_i = 12$ cm

16.2 Refraction

34. At what minimum angle does total internal reflection of light occur if it travels from water $(n = 1.33)$ toward ice $(n = 1.31)$?

a. $44.6°$
b. $26.5°$
c. $13.3°$
d. $80.1°$

35. Water floats on a liquid called *carbon tetrachloride*. The two liquids do not mix. A light ray passing from water into carbon tetrachloride has an incident angle of 45.0° and an angle of refraction of 40.1°. If the index of refraction of water is 1.33, what is the index of refraction of carbon tetrachloride?

a. 1.60
b. 1.49
c. 1.21
d. 1.46

36. Describe what happens to a light ray when it is refracted. Include in your explanation comparison of angles, comparison of refractive indices, and the term *normal*.

a. When a ray of light goes from one medium to another medium with a different refractive index, the ray changes its path as a result of interference. The angle between the ray and the normal (the line perpendicular to the surfaces of the two media) is greater in the medium with the greater refractive index.
b. When a ray of light goes from one medium to another medium with a different refractive index, the ray changes its path as a result of refraction. The angle between the ray and the normal (the line perpendicular to the surfaces of the two media) is less in the medium with the greater refractive index.
c. When a ray of light goes from one medium to another medium with a different refractive index, the ray does not change its path. The angle between the ray and the normal (the line parallel to the surfaces of the two media) is the same in both media.
d. When a ray of light goes from one medium to another medium with a different refractive index, the ray changes its path as a result of refraction. The angle between the ray and the normal (the line perpendicular to the surfaces of the two media) is less in the medium with the lower refractive index.

16.3 Lenses

37. What are two equivalent terms for a lens that always causes light rays to bend away from the principal axis?

a. a diverging lens or a convex lens
b. a diverging lens or a concave lens
c. a converging lens or a concave lens
d. a converging lens or a convex lens

38. Define the term *virtual image*.

a. A virtual image is an image that cannot be projected onto a screen.
b. A virtual image is an image that can be projected onto a screen.
c. A virtual image is an image that is formed on the opposite side of the lens from where the object is placed.
d. A virtual image is an image that is always bigger than the object.

39. Compare nearsightedness (myopia) and farsightedness (hyperopia) in terms of focal point.

a. The eyes of a nearsighted person have focal points beyond the retina. A farsighted person has eyes with focal points between the lens and the retina.
b. A nearsighted person has eyes with focal points between the lens and the retina. A farsighted person has eyes with focal points beyond the retina.
c. A nearsighted person has eyes with focal points between the lens and the choroid. A farsighted person has eyes with focal points beyond the choroid.
d. A nearsighted person has eyes with focal points between the lens and the retina. A farsighted person has eyes with focal points on the retina.

40. Explain how a converging lens corrects farsightedness.

a. A converging lens disperses the rays so they focus on the retina.
b. A converging lens bends the rays closer together so they do not focus on the retina.
c. A converging lens bends the rays closer together so they focus on the retina.

d. A converging lens disperses the rays so they do not focus on the retina.

41. Solve the equation $\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f}$ for f in such a way that it is not expressed as a reciprocal.

a. $f = \frac{d_o + d_i}{d_o d_i}$
b. $f = \frac{d_o d_i}{d_i + d_o}$
c. $f = (d_i + d_o)$
d. $f = d_o d_i$

Extended Response

16.1 Reflection

43. The diagram shows a lightbulb between two mirrors. One mirror produces a beam of light with parallel rays; the other keeps light from escaping without being put into the beam.

Where is the light source in relation to the focal point or radius of curvature of each mirror? Explain your answer.

a. The bulb is at the center of curvature of the small mirror and at the focal point of the large mirror.
b. The bulb is at the focal point of the small mirror and at the focal point of the large mirror.
c. The bulb is at the center of curvature of the small mirror and at the center of curvature of the large mirror.
d. The bulb is at the focal point of the small mirror and at the center of curvature of the large mirror.

44. An object is placed $4.00\ \text{cm}$ in front of a mirror that has a magnification of 1.50. What is the radius of curvature of the mirror?

a. $-24.0\ \text{cm}$
b. $-4.80\ \text{cm}$
c. $4.80\ \text{cm}$
d. $24.0\ \text{cm}$

16.2 Refraction

45. A scuba diver training in a pool looks at his instructor, as shown in this figure. The angle between the ray in the water and the normal to the water is 25°.

What angle does the ray make from the instructor's face with the normal to the water ($n = 1.33$) at the point where the ray enters? Assume $n = 1.00$ for air.

a. 68°
b. 25°
c. 19°
d. 34°

46. Describe total internal reflection. Include a definition of the critical angle and how it is related to total internal reflection. Also, compare the indices of refraction of the interior material and the surrounding material.

a. When the interior material has a smaller index of refraction than the surrounding material, the incident ray may approach the boundary at an angle (called the *critical angle*) such that the refraction angle is 90° . The refracted ray cannot leave the interior, so it is reflected back inside and total internal reflection occurs.
b. When the interior material has a smaller index of refraction than the surrounding material, the incident ray may approach the boundary at an angle (called the *critical angle*) such that the refraction angle is less than 90° . The refracted ray cannot leave the interior, so it is reflected back inside and total internal reflection occurs.
c. When the interior material has the same index of refraction as the surrounding material, the incident ray approaches the boundary at an angle (called the *critical angle*) such that the refraction

angle is less than $90°$. The refracted ray cannot leave the interior, so it is reflected back inside and total internal reflection occurs.

d. When the interior material has a greater index of refraction than the surrounding material, the incident ray may approach the boundary at an angle (called the *critical angle*) such that the refraction angle is $90°$. The refracted ray cannot leave the interior, so it is reflected back inside and total internal reflection occurs.

16.3 Lenses

47. The muscles that change the shape of the lens in the eyes have become weak, causing vision problems for a person. In particular, the muscles cannot pull hard enough on the edges of the lens to make it less convex.
Part A—What condition does inability cause?
Part B—Where are images focused with respect to the retina?
Part C—Which type of lens corrects this person's problem? Explain.

a. Part A—This condition causes hyperopia.
Part B—Images are focused between the lens and the retina.
Part C—A converging lens gathers the rays slightly so they focus onto the retina.

b. Part A—This condition causes myopia.
Part B—Images are focused between the lens and the retina.
Part C—A converging lens gathers the rays slightly so they focus onto the retina.

c. Part A—This condition causes hyperopia.
Part B—Images are focused between the lens and the retina.
Part C—A diverging lens spreads the rays slightly so they focus onto the retina.

d. Part A—This condition causes myopia.
Part B—Images are focused between the lens and the retina.
Part C—A diverging lens spreads the rays slightly so they focus onto the retina.

48. If the lens-to-retina distance is $2.00\ \text{cm}$, what is the power of the eye when viewing an object $50.0\ \text{cm}$ away?

a. $-52.0\ \text{D}$
b. $0.52\ \text{D}$
c. $1.92\ \text{D}$
d. $52.0\ \text{D}$

CHAPTER 17
Diffraction and Interference

Figure 17.1 The colors reflected by this compact disc vary with angle and are not caused by pigments. Colors such as these are direct evidence of the wave character of light. (credit: Reggie Mathalone)

Chapter Outline

17.1 Understanding Diffraction and Interference

17.2 Applications of Diffraction, Interference, and Coherence

INTRODUCTION Examine a compact disc under white light, noting the colors observed and their locations on the disc. Using the CD, explore the spectra of a few light sources, such as a candle flame, an incandescent bulb, and fluorescent light. If you have ever looked at the reds, blues, and greens in a sunlit soap bubble and wondered how straw-colored soapy water could produce them, you have hit upon one of the many phenomena that can only be explained by the wave character of light. That and other interesting phenomena, such as the dispersion of white light into a rainbow of colors when passed through a narrow slit, cannot be explained fully by geometric optics. In such cases, light interacts with small objects and exhibits its wave characteristics. The topic of this chapter is the branch of optics that considers the behavior of light when it exhibits wave characteristics.

17.1 Understanding Diffraction and Interference

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Explain wave behavior of light, including diffraction and interference, including the role of constructive and destructive interference in Young's single-slit and double-slit experiments
- Perform calculations involving diffraction and interference, in particular the wavelength of light using data from a two-slit interference pattern

Section Key Terms

diffraction	Huygens's principle	monochromatic	wavefront

Diffraction and Interference

We know that visible light is the type of electromagnetic wave to which our eyes responds. As we have seen previously, light obeys the equation

$$c = f\lambda,$$

where $c = 3.00 \times 10^8$ m/s is the speed of light in vacuum, f is the frequency of the electromagnetic wave in Hz (or s^{-1}), and λ is its wavelength in m. The range of visible wavelengths is approximately 380 to 750 nm. As is true for all waves, light travels in straight lines and acts like a ray when it interacts with objects several times as large as its wavelength. However, when it interacts with smaller objects, it displays its wave characteristics prominently. Interference is the identifying behavior of a wave.

In Figure 17.2, both the ray and wave characteristics of light can be seen. The laser beam emitted by the observatory represents ray behavior, as it travels in a straight line. Passing a pure, one-wavelength beam through vertical slits with a width close to the wavelength of the beam reveals the wave character of light. Here we see the beam spreading out horizontally into a pattern of bright and dark regions that are caused by systematic constructive and destructive interference. As it is characteristic of wave behavior, interference is observed for water waves, sound waves, and light waves.

Figure 17.2 (a) The light beam emitted by a laser at the Paranal Observatory (part of the European Southern Observatory in Chile) acts like a ray, traveling in a straight line. (credit: Yuri Beletsky, European Southern Observatory) (b) A laser beam passing through a grid of vertical slits produces an interference pattern—characteristic of a wave. (credit: Shim'on and Slava Rybka, Wikimedia Commons)

That interference is a characteristic of energy propagation by waves is demonstrated more convincingly by water waves. Figure 17.3 shows water waves passing through gaps between some rocks. You can easily see that the gaps are similar in width to the wavelength of the waves and that this causes an interference pattern as the waves pass beyond the gaps. A cross-section across the waves in the foreground would show the crests and troughs characteristic of an interference pattern.

Figure 17.3 Incoming waves (at the top of the picture) pass through the gaps in the rocks and create an interference pattern (in the foreground).

Light has wave characteristics in various media as well as in a vacuum. When light goes from a vacuum to some medium, such as water, its speed and wavelength change, but its frequency, f, remains the same. The speed of light in a medium is $v = c/n$, where n is its index of refraction. If you divide both sides of the equation $c = f\lambda$ by n, you get $c/n = v = f\lambda/n$. Therefore, $v = f\lambda_n$, where λ_n is the wavelength in a medium, and

$$\lambda_n = \frac{\lambda}{n},$$

where λ is the wavelength in vacuum and n is the medium's index of refraction. It follows that the wavelength of light is smaller in any medium than it is in vacuum. In water, for example, which has $n = 1.333$, the range of visible wavelengths is (380 nm)/1.333 to (760 nm)/1.333, or $\lambda_n = 285–570$ nm. Although wavelengths change while traveling from one medium to another, colors do not, since colors are associated with frequency.

The Dutch scientist Christiaan Huygens (1629–1695) developed a useful technique for determining in detail how and where waves propagate. He used **wavefronts**, which are the points on a wave's surface that share the same, constant phase (such as all the points that make up the crest of a water wave). **Huygens's principle** states, "Every point on a wavefront is a source of wavelets that spread out in the forward direction at the same speed as the wave itself. The new wavefront is a line tangent to all of the wavelets."

Figure 17.4 shows how Huygens's principle is applied. A wavefront is the long edge that moves; for example, the crest or the trough. Each point on the wavefront emits a semicircular wave that moves at the propagation speed v. These are drawn later at a time, t, so that they have moved a distance $s = vt$. The new wavefront is a line tangent to the wavelets and is where the wave is located at time t. Huygens's principle works for all types of waves, including water waves, sound waves, and light waves. It will be useful not only in describing how light waves propagate, but also in how they interfere.

Figure 17.4 Huygens's principle applied to a straight wavefront. Each point on the wavefront emits a semicircular wavelet that moves a distance $s = vt$. The new wavefront is a line tangent to the wavelets.

What happens when a wave passes through an opening, such as light shining through an open door into a dark room? For light, you expect to see a sharp shadow of the doorway on the floor of the room, and you expect no light to bend around corners into other parts of the room. When sound passes through a door, you hear it everywhere in the room and, thus, you understand that sound spreads out when passing through such an opening. What is the difference between the behavior of sound waves and light waves in this case? The answer is that the wavelengths that make up the light are very short, so that the light acts like a ray. Sound has wavelengths on the order of the size of the door, and so it bends around corners.

If light passes through smaller openings, often called slits, you can use Huygens's principle to show that light bends as sound does (see Figure 17.5). The bending of a wave around the edges of an opening or an obstacle is called **diffraction**. Diffraction is a

wave characteristic that occurs for all types of waves. If diffraction is observed for a phenomenon, it is evidence that the phenomenon is produced by waves. Thus, the horizontal diffraction of the laser beam after it passes through slits in Figure 17.2 is evidence that light has the properties of a wave.

Figure 17.5 Huygens's principle applied to a straight wavefront striking an opening. The edges of the wavefront bend after passing through the opening, a process called diffraction. The amount of bending is more extreme for a small opening, consistent with the fact that wave characteristics are most noticeable for interactions with objects about the same size as the wavelength.

Once again, water waves present a familiar example of a wave phenomenon that is easy to observe and understand, as shown in Figure 17.6.

Figure 17.6 Ocean waves pass through an opening in a reef, resulting in a diffraction pattern. Diffraction occurs because the opening is similar in width to the wavelength of the waves.

WATCH PHYSICS

Single-Slit Interference

This video works through the math needed to predict diffraction patterns that are caused by single-slit interference.

Click to view content (https://www.openstax.org/l/28slit)

Which values of m denote the location of destructive interference in a single-slit diffraction pattern?

a. whole integers, excluding zero
b. whole integers
c. real numbers excluding zero
d. real numbers

The fact that Huygens's principle worked was not considered enough evidence to prove that light is a wave. People were also reluctant to accept light's wave nature because it contradicted the ideas of Isaac Newton, who was still held in high esteem. The acceptance of the wave character of light came after 1801, when the English physicist and physician Thomas Young (1773–1829) did his now-classic double-slit experiment (see Figure 17.7).

Figure 17.7 Young's double-slit experiment. Here, light of a single wavelength passes through a pair of vertical slits and produces a diffraction pattern on the screen—numerous vertical light and dark lines that are spread out horizontally. Without diffraction and interference, the light would simply make two lines on the screen.

When light passes through narrow slits, it is diffracted into semicircular waves, as shown in Figure 17.8 (a). Pure constructive interference occurs where the waves line up crest to crest or trough to trough. Pure destructive interference occurs where they line up crest to trough. The light must fall on a screen and be scattered into our eyes for the pattern to be visible. An analogous pattern for water waves is shown in Figure 17.8 (b). Note that regions of constructive and destructive interference move out from the slits at well-defined angles to the original beam. Those angles depend on wavelength and the distance between the slits, as you will see below.

Figure 17.8 Double slits produce two sources of waves that interfere. (a) Light spreads out (diffracts) from each slit, because the slits are narrow. The waves overlap and interfere constructively (bright lines) and destructively (dark regions). You can only see the effect if the light falls onto a screen and is scattered into your eyes. (b) The double-slit interference pattern for water waves is nearly identical to that for light. Wave action is greatest in regions of constructive interference and least in regions of destructive interference. (c) When light that has passed through double slits falls on a screen, we see a pattern such as this.

Virtual Physics

Wave Interference

Click to view content (https://www.openstax.org/l/28interference)

This simulation demonstrates most of the wave phenomena discussed in this section. First, observe interference between two sources of electromagnetic radiation without adding slits. See how water waves, sound, and light all show interference patterns. Stay with light waves and use only one source. Create diffraction patterns with one slit and then with two. You may have to adjust slit width to see the pattern.

Visually compare the slit width to the wavelength. When do you get the best-defined diffraction pattern?

a. when the slit width is larger than the wavelength
b. when the slit width is smaller than the wavelength
c. when the slit width is comparable to the wavelength
d. when the slit width is infinite

Calculations Involving Diffraction and Interference

The fact that the wavelength of light of one color, or **monochromatic** light, can be calculated from its two-slit diffraction pattern in Young's experiments supports the conclusion that light has wave properties. To understand the basis of such calculations, consider how two waves travel from the slits to the screen. Each slit is a different distance from a given point on the screen. Thus different numbers of wavelengths fit into each path. Waves start out from the slits in phase (crest to crest), but they will end up out of phase (crest to trough) at the screen if the paths differ in length by half a wavelength, interfering destructively. If the paths differ by a whole wavelength, then the waves arrive in phase (crest to crest) at the screen, interfering constructively. More generally, if the paths taken by the two waves differ by any half-integral number of wavelengths ($\frac{1}{2}\lambda,\ \frac{3}{2}\lambda,\ \frac{5}{2}\lambda,$ etc.), then destructive interference occurs. Similarly, if the paths taken by the two waves differ by any integral number of wavelengths ($\lambda,\ 2\lambda,\ 3\lambda,$ etc.), then constructive interference occurs.

Figure 17.9 shows how to determine the path-length difference for waves traveling from two slits to a common point on a screen. If the screen is a large distance away compared with the distance between the slits, then the angle θ between the path and a line from the slits perpendicular to the screen (see the figure) is nearly the same for each path. That approximation and simple trigonometry show the length difference, ΔL, to be $d \sin \theta$, where d is the distance between the slits,

$$\Delta L = d \sin \theta.$$

To obtain constructive interference for a double slit, the path-length difference must be an integral multiple of the wavelength, or

$$d \sin \theta = m\lambda, \text{ for } m = 0, 1, -1, 2, -2, \ldots \text{ (constructive).}$$

Similarly, to obtain destructive interference for a double slit, the path-length difference must be a half-integral multiple of the wavelength, or

$$d \sin \theta = (m + ½)\lambda, \text{ for } m = 0, 1, -1, 2, -2, \ldots \text{ (destructive).}$$

The number m is the *order* of the interference. For example, $m = 4$ is fourth-order interference.

Figure 17.9 The paths from each slit to a common point on the screen differ by an amount $d \sin \theta$, assuming the distance to the screen is much greater than the distance between the slits (not to scale here).

Figure 17.10 shows how the intensity of the bands of constructive interference decreases with increasing angle.

Figure 17.10 The interference pattern for a double slit has an intensity that falls off with angle. The photograph shows multiple bright and dark lines, or fringes, formed by light passing through a double slit.

Light passing through a single slit forms a diffraction pattern somewhat different from that formed by double slits. Figure 17.11 shows a single-slit diffraction pattern. Note that the central maximum is larger than those on either side, and that the intensity decreases rapidly on either side.

Figure 17.11 (a) Single-slit diffraction pattern. Monochromatic light passing through a single slit produces a central maximum and many smaller and dimmer maxima on either side. The central maximum is six times higher than shown. (b) The drawing shows the bright central maximum and dimmer and thinner maxima on either side. (c) The location of the minima are shown in terms of λ and D.

The analysis of single-slit diffraction is illustrated in Figure 17.12. Assuming the screen is very far away compared with the size of the slit, rays heading toward a common destination are nearly parallel. That approximation allows a series of trigonometric operations that result in the equations for the minima produced by destructive interference.

$$D\sin\theta = m\lambda$$

or

$$\frac{Dy}{L} = m\lambda$$

When rays travel straight ahead, they remain in phase and a central maximum is obtained. However, when rays travel at an angle θ relative to the original direction of the beam, each ray travels a different distance to the screen, and they can arrive in or out of phase. Thus, a ray from the center travels a distance $\lambda/2$ farther than the ray from the top edge of the slit, they arrive out of phase, and they interfere destructively. Similarly, for every ray between the top and the center of the slit, there is a ray between the center and the bottom of the slit that travels a distance $\lambda/2$ farther to the common point on the screen, and so interferes destructively. Symmetrically, there will be another minimum at the same angle below the direct ray.

Figure 17.12 Equations for a single-slit diffraction pattern, where λ is the wavelength of light, D is the slit width, θ is the angle between a line from the slit to a minimum and a line perpendicular to the screen, L is the distance from the slit to the screen, y is the distance from the center of the pattern to the minimum, and m is a nonzero integer indicating the order of the minimum.

Below we summarize the equations needed for the calculations to follow.

The speed of light in a vacuum, c, the wavelength of the light, λ, and its frequency, f, are related as follows.

$$c = f\lambda$$

The wavelength of light in a medium, λ_n, compared to its wavelength in a vacuum, λ, is given by

$$\lambda_n = \frac{\lambda}{n}. \tag{17.1}$$

To calculate the positions of constructive interference for a double slit, the path-length difference must be an integral multiple, m, of the wavelength. λ

$$d\sin\theta = m\lambda, \text{ for } m = 0, 1, -1, 2, -2, \ldots \text{ (constructive)},$$

where d is the distance between the slits and θ is the angle between a line from the slits to the maximum and a line perpendicular to the barrier in which the slits are located. To calculate the positions of destructive interference for a double slit, the path-length difference must be a half-integral multiple of the wavelength:

$$d\sin\theta = (m + ½)\lambda, \text{ for } m = 0, 1, -1, 2, -2, \ldots \text{ (destructive)}.$$

For a single-slit diffraction pattern, the width of the slit, D, the distance of the first ($m = 1$) destructive interference minimum, y, the distance from the slit to the screen, L, and the wavelength, λ, are given by

$$\frac{Dy}{L} = \lambda.$$

Also, for single-slit diffraction,

$$D\sin\theta = m\lambda,$$

where θ is the angle between a line from the slit to the minimum and a line perpendicular to the screen, and m is the order of the minimum.

WORKED EXAMPLE

Two-Slit Interference

Suppose you pass light from a He-Ne laser through two slits separated by 0.0100 mm, and you find that the third bright line on a screen is formed at an angle of 10.95° relative to the incident beam. What is the wavelength of the light?

STRATEGY

The third bright line is due to third-order constructive interference, which means that $m = 3$. You are given $d = 0.0100$ mm and $\theta = 10.95°$. The wavelength can thus be found using the equation $d\sin\theta = m\lambda$ for constructive interference.

Solution

The equation is $d \sin\theta = m\lambda$. Solving for the wavelength, λ, gives

$$\lambda = \frac{d \sin\theta}{m}. \tag{17.2}$$

Substituting known values yields

$$\lambda = \frac{(0.0100\ \text{mm})(\sin\ 10.95°)}{3} = 6.33 \times 10^{-4}\ \text{mm} = 633\ \text{nm}. \tag{17.3}$$

Discussion

To three digits, 633 nm is the wavelength of light emitted by the common He-Ne laser. Not by coincidence, this red color is similar to that emitted by neon lights. More important, however, is the fact that interference patterns can be used to measure wavelength. Young did that for visible wavelengths. His analytical technique is still widely used to measure electromagnetic spectra. For a given order, the angle for constructive interference increases with λ, so spectra (measurements of intensity versus wavelength) can be obtained.

WORKED EXAMPLE

Single-Slit Diffraction

Visible light of wavelength 550 nm falls on a single slit and produces its second diffraction minimum at an angle of 45.0° relative to the incident direction of the light. What is the width of the slit?

STRATEGY

From the given information, and assuming the screen is far away from the slit, you can use the equation $D \sin\theta = m\lambda$ to find D.

Solution

Quantities given are $\lambda = 550$ nm, $m = 2$, and $\theta_2 = 45.0°$. Solving the equation $D \sin\theta = m\lambda$ for D and substituting known values gives

$$D = \frac{m\lambda}{\sin\theta} = \frac{2(550\ \text{nm})}{\sin 45.0°} = 1.56 \times 10^{-6}\ \text{m}. \tag{17.4}$$

Discussion

You see that the slit is narrow (it is only a few times greater than the wavelength of light). That is consistent with the fact that light must interact with an object comparable in size to its wavelength in order to exhibit significant wave effects, such as this single-slit diffraction pattern.

Practice Problems

1. Monochromatic light from a laser passes through two slits separated by $0.00500\ \text{mm}$. The third bright line on a screen is formed at an angle of $18.0°$ relative to the incident beam. What is the wavelength of the light?
   a. $51.5\ \text{nm}$
   b. $77.3\ \text{nm}$
   c. $515\ \text{nm}$
   d. $773\ \text{nm}$

2. What is the width of a single slit through which 610-nm orange light passes to form a first diffraction minimum at an angle of 30.0°?
   a. 0.863 μm
   b. 0.704 μm
   c. 0.610 μm
   d. 1.22 μm

Check Your Understanding

3. Which aspect of a beam of monochromatic light changes when it passes from a vacuum into water, and how does it change?
   a. The wavelength first decreases and then increases.
   b. The wavelength first increases and then decreases.
   c. The wavelength increases.
   d. The wavelength decreases.

4. Go outside in the sunlight and observe your shadow. It has fuzzy edges, even if you do not. Is this a diffraction effect? Explain.
   a. This is a diffraction effect. Your whole body acts as the origin for a new wavefront.
   b. This is a diffraction effect. Every point on the edge of your shadow acts as the origin for a new wavefront.
   c. This is a refraction effect. Your whole body acts as the origin for a new wavefront.
   d. This is a refraction effect. Every point on the edge of your shadow acts as the origin for a new wavefront.

5. Which aspect of monochromatic green light changes when it passes from a vacuum into diamond, and how does it change?
   a. The wavelength first decreases and then increases.
   b. The wavelength first increases and then decreases.
   c. The wavelength increases.
   d. The wavelength decreases.

17.2 Applications of Diffraction, Interference, and Coherence

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Explain behaviors of waves, including reflection, refraction, diffraction, interference, and coherence, and describe applications based on these behaviors
- Perform calculations related to applications based on wave properties of light

Section Key Terms

differential interference contrast (DIC)	diffraction grating	iridescence	laser
monochromator	Rayleigh criterion	resolution	

Wave-Based Applications of Light

In 1917, Albert Einstein was thinking about photons and excited atoms. He considered an atom excited by a certain amount of energy and what would happen if that atom were hit by a photon with the same amount of energy. He suggested that the atom would emit a photon with that amount of energy, and it would be accompanied by the original photon. The exciting part is that you would have *two* photons with the same energy *and* they would be in phase. Those photons could go on to hit other excited atoms, and soon you would have a stream of in-phase photons. Such a light stream is said to be *coherent*. Some four decades later, Einstein's idea found application in a process called, *light amplification by stimulated emission of radiation*. Take the first letters of all the words (except *by* and "of") and write them in order. You get the word ***laser*** (see (a)), which is the name of the device that produces such a beam of light.

Laser beams are directional, very intense, and narrow (only about 0.5 mm in diameter). These properties lead to a number of applications in industry and medicine. The following are just a few examples:

- This chapter began with a picture of a compact disc (see). Those audio and data-storage devices began replacing cassette tapes during the 1990s. CDs are read by interpreting variations in reflections of a laser beam from the surface.
- Some barcode scanners use a laser beam.
- Lasers are used in industry to cut steel and other metals.
- Lasers are bounced off reflectors that astronauts left on the Moon. The time it takes for the light to make the round trip can be used to make precise calculations of the Earth-Moon distance.
- Laser beams are used to produce holograms. The name hologram means *entire picture* (from the Greek *holo-*, as in

holistic), because the image is three-dimensional. A viewer can move around the image and see it from different perspectives. Holograms take advantage of the wave properties of light, as opposed to traditional photography which is based on geometric optics. A holographic image is produced by constructive and destructive interference of a split laser beam.

- One of the advantages of using a laser as a surgical tool is that it is accompanied by very little bleeding.
- Laser eye surgery has improved the vision of many people, without the need for corrective lenses. A laser beam is used to change the shape of the lens of the eye, thus changing its focal length.

Virtual Physics

Lasers

Click to view content (https://www.openstax.org/l/28lasers)

This animation allows you to examine the workings of a laser. First view the picture of a real laser. Change the energy of the incoming photons, and see if you can match it to an excitation level that will produce pairs of coherent photons. Change the excitation level and try to match it to the incoming photon energy.

In the animation there is only one excited atom. Is that the case for a real laser? Explain.

a. No, a laser would have two excited atoms.
b. No, a laser would have several million excited atoms.
c. Yes, a laser would have only one excited atom.
d. No, a laser would have on the order of 10^{23} excited atoms.

An interesting thing happens if you pass light through a large number of evenly-spaced parallel slits. Such an arrangement of slits is called a **diffraction grating**. An interference pattern is created that is very similar to the one formed by double-slit diffraction (see and). A diffraction grating can be manufactured by scratching glass with a sharp tool to form a number of precisely positioned parallel lines, which act like slits. Diffraction gratings work both for transmission of light, as in Figure 17.13, and for reflection of light, as on the butterfly wings or the Australian opal shown in Figure 17.14, or the CD pictured in the opening illustration of this chapter. In addition to their use as novelty items, diffraction gratings are commonly used for spectroscopic dispersion and analysis of light. What makes them particularly useful is the fact that they form a sharper pattern than do double slits. That is, their bright regions are narrower and brighter, while their dark regions are darker. Figure 17.15 shows idealized graphs demonstrating the sharper pattern. Natural diffraction gratings occur in the feathers of certain birds. Tiny, fingerlike structures in regular patterns act as reflection gratings, producing constructive interference that gives the feathers colors not solely due to their pigmentation. The effect is called **iridescence**.

Figure 17.13 A diffraction grating consists of a large number of evenly-spaced parallel slits. (a) Light passing through the grating is diffracted in a pattern similar to a double slit, with bright regions at various angles. (b) The pattern obtained for white light incident on a grating. The central maximum is white, and the higher-order maxima disperse white light into a rainbow of colors.

Figure 17.14 (a) This Australian opal and (b) the butterfly wings have rows of reflectors that act like reflection gratings, reflecting different colors at different angles. (credit: (a) Opals-On-Black.com, via Flickr (b) whologwhy, Flickr)

Figure 17.15 Idealized graphs of the intensity of light passing through a double slit (a) and a diffraction grating (b) for monochromatic light. Maxima can be produced at the same angles, but those for the diffraction grating are narrower, and hence sharper. The maxima become narrower and the regions between become darker as the number of slits is increased.

Snap Lab

Diffraction Grating

- A CD (compact disc) or DVD
- A measuring tape
- Sunlight near a white wall

Instructions

Procedure

1. Hold the CD in direct sunlight near the wall, and move it around until a circular rainbow pattern appears on the wall.
2. Measure the distance from the CD to the wall and the distance from the center of the circular pattern to a color in the rainbow. Use those two distances to calculate $\tan\theta$. Find $\sin\theta$.
3. Look up the wavelength of the color you chose. That is λ.
4. Solve $d\sin\theta = m\lambda$ for d.
5. Compare your answer to the usual spacing between CD tracks, which is 1,600 nm (1.6 µm).
   How do you know what number to use for m?
   a. Count the rainbow rings preceding the chosen color.

b. Calculate mfrom the frequency of the light of the chosen color.
c. Calculate m from the wavelength of the light of the chosen color.
d. The value of m is fixed for every color.

FUN IN PHYSICS

CD Players

Can you see the grooves on a CD or DVD (see Figure 17.16)? You may think you can because you know they are there, but they are extremely narrow—1,600 in a millimeter. Because the width of the grooves is similar to wavelengths of visible light, they form a diffraction grating. That is why you see rainbows on a CD. The colors are attractive, but they are incidental to the functions of storing and retrieving audio and other data.

Figure 17.16 For its size, this CD holds a surprising amount of information. Likewise, the CD player it is in houses a surprising number of electronic devices.

The grooves are actually one continuous groove that spirals outward from the center. Data are recorded in the grooves as binary code (zeroes and ones) in small pits. Information in the pits is detected by a laser that tracks along the groove. It gets even more complicated: The speed of rotation must be varied as the laser tracks toward the circumference so that the linear speed along the groove remains constant. There is also an error correction mechanism to prevent the laser beam from getting off track. A diffraction grating is used to create the first two maxima on either side of the track. If those maxima are not the same distance from the track, an error is indicated and then corrected.

The pits are reflective because they have been coated with a thin layer of aluminum. That allows the laser beam to be reflected back and directed toward a photodiode detector. The signal can then be processed and converted to the audio we hear.

The longest wavelength of visible light is about 780 nm . How does that compare to the distance between CD grooves?

a. The grooves are about 3 times the longest wavelength of visible light.
b. The grooves are about 2 times the longest wavelength of visible light.
c. The grooves are about 2 times the shortest wavelength of visible light.
d. The grooves are about 3 times the shortest wavelength of visible light.

LINKS TO PHYSICS

Biology: DIC Microscopy

If you were completely transparent, it would be hard to recognize you from your photograph. The same problem arises when using a traditional microscope to view or photograph small transparent objects such as cells and microbes. Microscopes using **differential interference contrast (DIC)** solve the problem by making it possible to view microscopic objects with enhanced contrast, as shown in Figure 17.17.

Figure 17.17 This aquatic organism was photographed with a DIC microscope. (credit: Public Library of Science)

A DIC microscope separates a polarized light source into two beams polarized at right angles to each other and coherent with each other, that is, in phase. After passing through the sample, the beams are recombined and realigned so they have the same plane of polarization. They then create an interference pattern caused by the differences in their optical path and the refractive indices of the parts of the sample they passed through. The result is an image with contrast and shadowing that could not be observed with traditional optics.

Where are diffraction gratings used? Diffraction gratings are key components of **monochromators**—devices that separate the various wavelengths of incoming light and allow a beam with only a specific wavelength to pass through. Monochromators are used, for example, in optical imaging of particular wavelengths from biological or medical samples. A diffraction grating can be chosen to specifically analyze a wavelength of light emitted by molecules in diseased cells in a biopsy sample, or to help excite strategic molecules in the sample with a selected frequency of light. Another important use is in optical fiber technologies where fibers are designed to provide optimum performance at specific wavelengths. A range of diffraction gratings is available for selecting specific wavelengths for such use.

Diffraction gratings are used in spectroscopes to separate a light source into its component wavelengths. When a material is heated to incandescence, it gives off wavelengths of light characteristic of the chemical makeup of the material. A pure substance will produce a spectrum that is unique, thus allowing identification of the substance. Spectroscopes are also used to measure wavelengths both shorter and longer than visible light. Such instruments have become especially useful to astronomers and chemists. Figure 17.18 shows a diagram of a spectroscope.

Figure 17.18 The diagram shows the function of a diffraction grating in a spectroscope.

Light diffracts as it moves through space, bending around obstacles and interfering constructively and destructively. While diffraction allows light to be used as a spectroscopic tool, it also limits the detail we can obtain in images.

Figure 17.19 (a) shows the effect of passing light through a small circular aperture. Instead of a bright spot with sharp edges, a spot with a fuzzy edge surrounded by circles of light is obtained. This pattern is caused by diffraction similar to that produced by a single slit. Light from different parts of the circular aperture interferes constructively and destructively. The effect is most noticeable when the aperture is small, but the effect is there for large apertures, too.

Figure 17.19 (a) Monochromatic light passed through a small circular aperture produces this diffraction pattern. (b) Two point light sources that are close to one another produce overlapping images because of diffraction. (c) If they are closer together, they cannot be resolved, that is, distinguished.

How does diffraction affect the detail that can be observed when light passes through an aperture? Figure 17.19 (b) shows the diffraction pattern produced by two point light sources that are close to one another. The pattern is similar to that for a single point source, and it is just barely possible to tell that there are two light sources rather than one. If they are closer together, as in Figure 17.19 (c), you cannot distinguish them, thus limiting the detail, or **resolution**, you can obtain. That limit is an inescapable consequence of the wave nature of light.

There are many situations in which diffraction limits the resolution. The acuity of vision is limited because light passes through the pupil, the circular aperture of the eye. Be aware that the diffraction-like spreading of light is due to the limited diameter of a light beam, not the interaction with an aperture. Thus light passing through a lens with a diameter of D shows the diffraction effect and spreads, blurring the image, just as light passing through an aperture of diameter D does. Diffraction limits the resolution of any system having a lens or mirror. Telescopes are also limited by diffraction, because of the finite diameter, D, of their primary mirror.

Why are diffraction gratings used in spectroscopes rather than just two slits?

a. The bands produced by diffraction gratings are dimmer but sharper than the bands produced by two slits.
b. The bands produced by diffraction gratings are brighter, though less sharp, than the bands produced by two slits.
c. The bands produced by diffraction gratings are brighter and sharper than the bands produced by two slits.
d. The bands produced by diffraction gratings are dimmer and less sharp, but more widely dispersed, than the bands produced by two slits.

Calculations Involving Diffraction Gratings and Resolution

Early in the chapter, it was mentioned that when light passes from one medium to another, its speed and wavelength change, but its frequency remains constant. The equation

$$\lambda_n = \frac{\lambda}{n}$$

shows how to the wavelength in a given medium, λ_n, is related to the wavelength in a vacuum, λ, and the refractive index, n, of the medium. The equation is useful for calculating the change in wavelength of a monochromatic laser beam in various media. The analysis of a diffraction grating is very similar to that for a double slit. As you know from the discussion of double slits in Young's double-slit experiment, light is diffracted by, and spreads out after passing through, each slit. Rays travel at an angle θ relative to the incident direction. Each ray travels a different distance to a common point on a screen far away. The rays start in phase, and they can be in or out of phase when they reach a screen, depending on the difference in the path lengths traveled. Each ray travels a distance that differs by $d \sin\theta$ from that of its neighbor, where d is the distance between slits. If $d \sin\theta$ equals an integral number of wavelengths, the rays all arrive in phase, and constructive interference (a maximum) is obtained. Thus, the condition necessary to obtain constructive interference for a diffraction grating is

$$d \sin\theta = m\lambda, \text{for } m = 0, 1, -1, 2, -2, \ldots,$$

where d is the distance between slits in the grating, λ is the wavelength of the light, and m is the order of the maximum. Note that this is exactly the same equation as for two slits separated by d. However, the slits are usually closer in diffraction gratings than in double slits, producing fewer maxima at larger angles.

WATCH PHYSICS

Diffraction Grating

This video (https://www.openstax.org/l/28diffraction) explains the geometry behind the diffraction pattern produced by a diffraction grating.

Click to view content (https://www.openstax.org/l/28diffraction)

The equation that gives the points of constructive interference produced by a diffraction grating is $d \sin\theta = m\lambda$. Why does that equation look familiar?

a. It is the same as the equation for destructive interference for a double-slit diffraction pattern.
b. It is the same as the equation for constructive interference for a double-slit diffraction pattern.
c. It is the same as the equation for constructive interference for a single-slit diffraction pattern.
d. It is the same as the equation for destructive interference for a single-slit diffraction pattern.

Just what is the resolution limit of an aperture or lens? To answer that question, consider the diffraction pattern for a circular aperture, which, similar to the diffraction pattern of light passing through a slit, has a central maximum that is wider and brighter than the maxima surrounding it (see Figure 17.19 (a)). It can be shown that, for a circular aperture of diameter D, the first minimum in the diffraction pattern occurs at $\theta = 1.22\lambda/D$, provided that the aperture is large compared with the wavelength of light, which is the case for most optical instruments. The accepted criterion for determining the diffraction limit to resolution based on diffraction was developed by Lord Rayleigh in the 19th century. The **Rayleigh criterion** for the diffraction limit to resolution states that two images are just resolvable when the center of the diffraction pattern of one is directly over the first minimum of the diffraction pattern of the other. See Figure 17.20 (b). The first minimum is at an angle of $\theta = 1.22\lambda/D$, so that two point objects are just resolvable if they are separated by the angle

$$\theta = 1.22\frac{\lambda}{D},$$

where λ is the wavelength of the light (or other electromagnetic radiation) and D is the diameter of the aperture, lens, mirror, etc., with which the two objects are observed. In the expression above, θ has units of radians.

Figure 17.20 (a) Graph of intensity of the diffraction pattern for a circular aperture. Note that, similar to a single slit, the central maximum is wider and brighter than those to the sides. (b) Two point objects produce overlapping diffraction patterns. Shown here is the Rayleigh criterion for their being just resolvable. The central maximum of one pattern lies on the first minimum of the other.

Snap Lab

Resolution

- A sheet of white paper
- A black pen or pencil
- A measuring tape

Instructions

Procedure

1. Draw two lines several mm apart on a white sheet of paper.
2. Move away from the sheet as it is held upright, and measure the distance at which you can just distinguish (resolve) the lines as separate.
3. Use $\theta = 1.22\frac{\lambda}{D}$ to calculate D the diameter of your pupil. Use the distance between the lines and the maximum distance at which they were resolved to calculate θ . Use the average wavelength for visible light as the value for λ .
4. Compare your answer to the average pupil diameter of 3 mm.

Describe resolution in terms of minima and maxima of diffraction patterns.

a. The limit for resolution is when the minimum of the pattern for one of the lines is directly over the first minimum of the pattern for the other line.
b. The limit for resolution is when the maximum of the pattern for one of the lines is directly over the first minimum of the pattern for the other line.
c. The limit for resolution is when the maximum of the pattern for one of the lines is directly over the second minimum of the pattern for the other line.
d. The limit for resolution is when the minimum of the pattern for one of the lines is directly over the second maximum of the pattern for the other line.

WORKED EXAMPLE

Change of Wavelength

A monochromatic laser beam of green light with a wavelength of 550 nm passes from air to water. The refractive index of water is 1.33. What will be the wavelength of the light after it enters the water?

STRATEGY

You can assume that the refractive index of air is the same as that of light in a vacuum because they are so close. You then have all the information you need to solve for λ_n .

Solution

$$\lambda_n = \frac{\lambda}{n} = \frac{550\ \text{nm}}{1.33} = 414\ \text{nm} \tag{17.5}$$

Discussion

The refractive index of air is 1.0003, so the approximation holds for three significant figures. You would not see the light change color, however. Color is determined by frequency, not wavelength.

WORKED EXAMPLE

Diffraction Grating

A diffraction grating has 2000 lines per centimeter. At what angle will the first-order maximum form for green light with a wavelength of 520 nm?

STRATEGY

You are given enough information to calculate d, and you are given the values of λ and m. You will have to find the arcsin of a

number to find θ.

Solution

First find d.

$$d = \frac{1 \text{ cm}}{2{,}000} = 5.00 \times 10^{-4} \text{ cm} = 5{,}000 \text{ nm} \tag{17.6}$$

Rearrange the equation for constructive interference conditions for a diffraction grating, and substitute the known values.

$$\begin{aligned} d \sin\theta &= m\lambda \\ \theta &= \sin^{-1} \tfrac{m\lambda}{d} \\ &= \sin^{-1}\left(\tfrac{(1)(520)}{5{,}000}\right) \\ &= 5.97 \end{aligned}$$

Discussion

This angle seems reasonable for the first maximum. Recall that the meaning of $\sin^{-1}$ (or arcsin) is *the angle with a sine that is (the unknown).* Remember that the value of $\sin\theta$ will not be greater than 1 for any value of θ.

WORKED EXAMPLE

Resolution

What is the minimum angular spread of a 633-nm-wavelength He-Ne laser beam that is originally 1.00 mm in diameter?

STRATEGY

The diameter of the beam is the same as if it were coming through an aperture of that size, so $D = 1.00$ mm. You are given λ, and you must solve for θ.

Solution

$$\theta = \frac{(1.22)\lambda}{D} = \frac{(1.22)(633 \text{ nm})}{1.00 \times 10^{6} \text{ nm}} = 7.72 \times 10^{-4} \text{ rad} = 0.0442^\circ \tag{17.7}$$

Discussion

The conversion factor for radians to degrees is 1.000 radian = 57.3°. The spread is very small and would not be noticeable over short distances. The angle represents the angular separation of the central maximum and the first minimum.

Practice Problems

6. A beam of yellow light has a wavelength of 600 nm in a vacuum and a wavelength of 397 nm in Plexiglas. What is the refractive index of Plexiglas?
   a. 1.51
   b. 2.61
   c. 3.02
   d. 3.77

7. What is the angle between two just-resolved points of light for a 3.00 mm diameter pupil, assuming an average wavelength of 550 nm?
   a. 224 rad
   b. 183 rad
   c. 1.83×10^{-4} rad
   d. 2.24×10^{-4} rad

Check Your Understanding

8. How is an interference pattern formed by a diffraction grating different from the pattern formed by a double slit?

a. The pattern is colorful.
b. The pattern is faded.
c. The pattern is sharper.
d. The pattern is curved.

9. A beam of light always spreads out. Why can a beam not be produced with parallel rays to prevent spreading?
a. Light is always polarized.
b. Light is always reflected.
c. Light is always refracted.
d. Light is always diffracted.

10. Compare interference patterns formed by a double slit and by a diffraction grating in terms of brightness and narrowness of bands.
a. The pattern formed has broader and brighter bands.
b. The pattern formed has broader and duller bands.
c. The pattern formed has narrower and duller bands.
d. The pattern formed has narrower and brighter bands.

11. Describe the slits in a diffraction grating in terms of number and spacing, as compared to a two-slit diffraction setup.
a. The slits in a diffraction grating are broader, with space between them that is greater than the separation of the two slits in two-slit diffraction.
b. The slits in a diffraction grating are broader, with space between them that is the same as the separation of the two slits in two-slit diffraction.
c. The slits in a diffraction grating are narrower, with space between them that is the same as the separation of the two slits in two-slit diffraction.
d. The slits in a diffraction grating are narrower, with space between them that is greater than the separation of the two slits in two-slit diffraction.

KEY TERMS

differential interference contrast (DIC) separating a polarized light source into two beams polarized at right angles to each other and coherent with each other then, after passing through the sample, recombining and realigning the beams so they have the same plane of polarization, and then creating an interference pattern caused by the differences in their optical path and the refractive indices of the parts of the sample they passed through; the result is an image with contrast and shadowing that could not be observed with traditional optics

diffraction bending of a wave around the edges of an opening or an obstacle

diffraction grating many of evenly spaced slits having dimensions such that they produce an interference pattern

Huygens's principle Every point on a wavefront is a source of wavelets that spread out in the forward direction at the same speed as the wave itself; the new wavefront is a line tangent to all of the wavelets.

iridescence the effect that occurs when tiny, fingerlike structures in regular patterns act as reflection gratings, producing constructive interference that gives feathers colors not solely due to their pigmentation

laser acronym for a device that produces *light amplification by stimulated emission of radiation*

monochromatic one color

monochromator device that separates the various wavelengths of incoming light and allows a beam with only a specific wavelength to pass through

Rayleigh criterion two images are just resolvable when the center of the diffraction pattern of one is directly over the first minimum of the diffraction pattern of the other

resolution degree to which two images can be distinguished from one another, which is limited by diffraction

wavefront points on a wave surface that all share an identical, constant phase

SECTION SUMMARY

17.1 Understanding Diffraction and Interference

- The wavelength of light varies with the refractive index of the medium.
- Slits produce a diffraction pattern if their width and separation are similar to the wavelength of light passing through them.
- Interference bands of a single-slit diffraction pattern can be predicted.
- Interference bands of a double-slit diffraction pattern can be predicted.

17.2 Applications of Diffraction, Interference, and Coherence

- The focused, coherent radiation emitted by lasers has many uses in medicine and industry.
- Characteristics of diffraction patterns produced with diffraction gratings can be determined.
- Diffraction gratings have been incorporated in many instruments, including microscopes and spectrometers.
- Resolution has a limit that can be predicted.

KEY EQUATIONS

17.1 Understanding Diffraction and Interference

speed of light, frequency, and wavelength	$c = f\lambda$
change of wavelength with index of refraction	$\lambda_n = \frac{\lambda}{n}$
two-slit constructive interference	$d \sin\theta = m\lambda$, for $m = 0, 1, -1, 2, -2, \ldots$
two-slit destructive interference	$d \sin\theta = (m + ½)\lambda$, for $m = 0, 1, -1, 2, -2, \ldots$
one-slit, first-order destructive interference; wavelength related to dimensions	$\frac{Dy}{L} = \lambda$
one-slit destructive interference	$D \sin\theta = m\lambda$

17.2 Applications of Diffraction, Interference, and Coherence

wavelength change with change in medium	$\lambda_n = \frac{\lambda}{n}$
diffraction grating constructive interference	$d \sin\theta = m\lambda$
resolution	$\theta = 1.22\frac{\lambda}{D}$

CHAPTER REVIEW

Concept Items

17.1 Understanding Diffraction and Interference

1. Which behavior of light is indicated by an interference pattern?
   a. ray behavior
   b. particle behavior
   c. corpuscular behavior
   d. wave behavior

2. Which behavior of light is indicated by diffraction?
   a. wave behavior
   b. particle behavior
   c. ray behavior
   d. corpuscular behavior

17.2 Applications of Diffraction, Interference, and Coherence

3. There is a principle related to resolution that is expressed by this equation.

$$\theta = \frac{\lambda}{D} \qquad 17.8$$

What is that principle stated in full?

4. A principle related to resolution states, "Two images are just resolved when the center of the diffraction pattern of one is directly over the first minimum of the diffraction pattern of the other." Write the equation that expresses that principle.
   a. $\theta = 1.22\frac{D}{\lambda}$
   b. $\theta = \frac{D}{\lambda}$
   c. $\theta = \frac{\lambda}{D}$
   d. $\theta = 1.22\frac{\lambda}{D}$

5. Which statement completes this resolution?
   Two images are just resolved when —
   a. The center of the diffraction pattern of one image is directly over the central maximum of the diffraction pattern of the other.
   b. The center of the diffraction pattern of one image is directly over the central minimum of the diffraction pattern of the other
   c. The center of the diffraction pattern of one image is directly over the first minimum of the diffraction pattern of the other
   d. The center of the diffraction pattern of one is directly over the first maximum of the diffraction pattern of the other

Critical Thinking Items

17.1 Understanding Diffraction and Interference

6. Describe a situation in which bodies of water and a line of rocks could create a diffraction pattern similar to light passing through double slits. Include the arrangement of the rocks, the positions of the bodies of water, and the location of the diffraction pattern. Note the dimensions that are necessary for the production of the pattern.
   a. When waves from a small body of water pass through two widely separated openings and enter a larger body of water, a diffraction pattern is produced that is similar tothe diffraction pattern formed by light passing through two slits. The width of each opening is larger than the size of the wavelength of the waves.
   b. When waves from a large body of water pass through two narrow openings and enter a smaller body of water, a diffraction pattern is produced that is similar to the diffraction pattern formed by light passing through two slits. The widths and separation of the openings are similar to the size of the wavelength of the waves.
   c. When waves from a small body of water pass through two wide openings and enter a larger body of water, a diffraction pattern is produced that is similar tothe diffraction pattern formed by light passing through two slits. The separation between the openings is similar to the size of the wavelength of the waves.
   d. When waves from a large body of water pass through two wide openings and enter a smaller body of water, a diffraction pattern is produced that is similar to the diffraction pattern formed by light passing through two slits. The widths and

separation of the openings are larger than the size of the wavelength of the waves.

17.2 Applications of Diffraction, Interference, and Coherence

7. For what type of electromagnetic radiation would a grating with spacing greater than 800 nm be useful as a spectroscopic tool?
   a. It can be used to analyze spectra only in the infrared portion of the spectrum.
   b. It can be used to analyze spectra in the entire visible portion of the electromagnetic spectrum.
   c. It can only be used to analyze spectra in the short-wavelength visible.
   d. It can only be used to analyze spectra in the short-wavelength visible and ultraviolet.

8. A beam of green light has a wavelength of 500 nm in a vacuum and a wavelength of 331 nm in Plexiglas. What is the refractive index of Plexiglas?
   a. 1.12
   b. 1.25
   c. 1.51
   d. 4.53

Problems

17.1 Understanding Diffraction and Interference

9. What is the distance between two slits that produce a diffraction pattern with the first minimum at an angle of 45.0° when 410-nm violet light passes through the slits?
   a. 2,030 nm
   b. 1,450 nm
   c. 410 nm
   d. 290 nm

10. A breakwater at the entrance to a harbor consists of a rock barrier with a 50.0 – m -wide opening. Ocean waves with a 20.0-m wavelength approach the opening straight on. At what angle to the incident direction are the boats inside the harbor most protected against wave action?
   a. 11.5°
   b. 7.46°
   c. 5.74°
   d. 23.6°

17.2 Applications of Diffraction, Interference, and Coherence

11. A 500-nm beam of light passing through a diffraction grating creates its second band of constructive interference at an angle of 1.50°. How far apart are the slits in the grating?
   a. 38,200 nm
   b. 19,100 nm
   c. 667 nm
   d. 333 nm

12. The range of the visible-light spectrum is 380 nm to 780 nm. What is the maximum number of lines per centimeter a diffraction grating can have and produce a complete first-order spectrum for visible light?
   a. 26,300 lines/cm
   b. 13,200 lines/cm
   c. 6,410 lines/cm
   d. 12,820 lines/cm

Performance Task

17.2 Applications of Diffraction, Interference, and Coherence

13. In this performance task you will create one- and two-slit diffraction and observe the interference patterns that result.
   - A utility knife (a knife with a razor blade-like cutting edge)
   - Aluminum foil
   - A straight edge
   - A strong, small light source or a laser pointer
   - A tape measure
   - A white wall

Procedure

1. Cut a piece of aluminum foil about 15 cm × 15 cm.
2. Use the utility knife and the straight edge to cut a straight slit several cm long in the center of the foil square.
3. With the room darkened, one partner shines the light through the slit and toward the wall. The other partner observes the pattern on the wall. The partner with the light changes the distance from the foil to the wall and the distance from the light to the foil.
4. When the sharpest, brightest pattern possible is obtained, the partner who is not holding the foil and light makes measurements.
5. Measure the perpendicular (shortest) distance from

the slit to the wall, the distance from the center of the pattern to several of the dark bands, and the distance from the slit to the same dark bands.

6. Carefully make a second slit parallel to the first slit and 1 mm or less away.
7. Repeat steps 2 through 5, only this time measure the distances to bright bands.
   NOTE—In your calculations, use 580 nm for λ if you used white light. If you used a colored laser pointer, look up the wavelength of the color. You may find it easier to calculate θ from its tangent rather than from its sine.
   a. Which experiment gave the most distinct pattern—one or two slits?
   b. What was the width of the single slit? Compare the calculated distance with the measured distance.
   c. What was the distance between the two slits? Compare the calculated distance with the measured distance.

TEST PREP

Multiple Choice

17.1 Understanding Diffraction and Interference

14. Which remains unchanged when a monochromatic beam of light passes from air into water?
   a. the speed of the light
   b. the direction of the beam
   c. the frequency of the light
   d. the wavelength of the light

15. Two slits are separated by a distance of 3500 nm . If light with a wavelength of 500 nm passes through the slits and produces an interference pattern, the m = _____ order minimum appears at an angle of 30.0°.
   a. 0
   b. 1
   c. 2
   d. 3

16. In the sunlight, the shadow of a building has fuzzy edges even if the building does not. Is this a refraction effect? Explain.
   a. Yes, this is a refraction effect, where every point on the building acts as the origin for a new wavefront.
   b. Yes, this is a refraction effect, where the whole building acts as the origin for a new wavefront.
   c. No, this is a diffraction effect, where every point on the edge of the building's shadow acts as the origin for a new wavefront.
   d. No, this is a diffraction effect, where the whole building acts as the origin for a new wavefront.

17.2 Applications of Diffraction, Interference, and Coherence

17. Two images are just resolved when the center of the diffraction pattern of one is directly over _____ of the diffraction pattern of the other.
   a. the center
   b. the first minimum
   c. the first maximum
   d. the last maximum

18. Two point sources of $500\ \text{nm}$ light are just resolvable as they pass through a small hole. The angle to the first minimum of one source is $0.100\ \text{rad}$. What is the diameter of the hole?
   a. $4.10\ \mu\text{m}$
   b. $5.73\ \mu\text{m}$
   c. $0.106\ \mu\text{m}$
   d. $6.10\ \mu\text{m}$

19. Will a beam of light shining through a 1-mm hole behave any differently than a beam of light that is 1 mm wide as it leaves its source? Explain.?
   a. Yes, the beam passing through the hole will spread out as it travels, because it is diffracted by the edges of the hole, whereas the 1 -mm beam, which encounters no diffracting obstacle, will not spread out.
   b. Yes, the beam passing through the hole will be made *more parallel* by passing through the hole, and so will not spread out as it travels, whereas the unaltered wavefronts of the 1-mm beam will cause the beam to spread out as it travels.
   c. No, both beams will remain the same width as they travel, and they will not spread out.
   d. No, both beams will spread out as they travel.

20. A laser pointer emits a coherent beam of parallel light rays. Does the light from such a source spread out at all? Explain.
   a. Yes, every point on a wavefront is not a source of wavelets, which prevent the spreading of light waves.
   b. No, every point on a wavefront is not a source of wavelets, so that the beam behaves as a bundles of rays that travel in their initial direction.
   c. No, every point on a wavefront is a source of

wavelets, which keep the beam from spreading.

d. Yes, every point on a wavefront is a source of wavelets, which cause the beam to spread out steadily as it moves forward.

Short Answer

17.1 Understanding Diffraction and Interference

21. Light passing through double slits creates a diffraction pattern. How would the spacing of the bands in the pattern change if the slits were closer together?

a. The bands would be closer together.
b. The bands would spread farther apart.
c. The bands would remain stationary.
d. The bands would fade and eventually disappear.

22. A beam of light passes through a single slit to create a diffraction pattern. How will the spacing of the bands in the pattern change if the width of the slit is increased?

a. The width of the spaces between the bands will remain the same.
b. The width of the spaces between the bands will increase.
c. The width of the spaces between the bands will decrease.
d. The width of the spaces between the bands will first decrease and then increase.

23. What is the wavelength of light falling on double slits separated by $2.00\ \mu\text{m}$ if the third-order maximum is at an angle of $60.0°$?

a. $667\ \text{nm}$
b. $471\ \text{nm}$
c. $333\ \text{nm}$
d. $577\ \text{nm}$

24. What is the longest wavelength of light passing through a single slit of width 1.20 μm for which there is a first-order minimum?

a. 1.04 μm
b. 0.849 μm
c. 0.600 μm
d. 2.40 μm

17.2 Applications of Diffraction, Interference, and Coherence

25. Describe a diffraction grating and the interference pattern it produces.

a. A diffraction grating is a large collection of evenly spaced parallel lines that produces an interference pattern that is similar to but sharper and better dispersed than that of a double slit.
b. A diffraction grating is a large collection of randomly spaced parallel lines that produces an interference pattern that is similar to but less sharp or well-dispersed as that of a double slit.
c. A diffraction grating is a large collection of randomly spaced intersecting lines that produces an interference pattern that is similar to but sharper and better dispersed than that of a double slit.
d. A diffraction grating is a large collection of evenly spaced intersecting lines that produces an interference pattern that is similar to but less sharp or well-dispersed as that of a double slit.

26. Suppose pure-wavelength light falls on a diffraction grating. What happens to the interference pattern if the same light falls on a grating that has more lines per centimeter?

a. The bands will spread farther from the central maximum.
b. The bands will come closer to the central maximum.
c. The bands will not spread farther from the first maximum.
d. The bands will come closer to the first maximum.

27. How many lines per centimeter are there on a diffraction grating that gives a first-order maximum for 473 nm blue light at an angle of 25.0°?

a. 529,000 lines/cm
b. 50,000 lines/cm
c. 851 lines/cm
d. 8,934 lines/cm

28. What is the distance between lines on a diffraction grating that produces a second-order maximum for 760-nm red light at an angle of 60.0°?

a. 2.28×10^4 nm
b. 3.29×10^2 nm
c. 2.53×10^1 nm
d. 1.76×10^3 nm

Extended Response

17.1 Understanding Diffraction and Interference

29. Suppose you use a double slit to perform Young's double-slit experiment in air, and then repeat the experiment with the same double slit in water. Does the color of the light change? Do the angles to the same parts of the interference pattern get larger or smaller? Explain.

a. No, the color is determined by frequency. The magnitude of the angle decreases.
b. No, the color is determined by wavelength. The magnitude of the angle decreases.
c. Yes, the color is determined by frequency. The magnitude of the angle increases.
d. Yes, the color is determined by wavelength. The magnitude of the angle increases.

30. A double slit is located at a distance x from a screen, with the distance along the screen from the center given by y. When the distance d between the slits is relatively large, there will be numerous bright bands.
For small angles (where $\sin\theta = \theta$, with θ in radians), what is the distance between fringes?

a. $\Delta y = \frac{d}{x\lambda}$
b. $\Delta y = \frac{xd}{\lambda}$
c. $\Delta y = \frac{\lambda}{xd}$
d. $\Delta y = \frac{x\lambda}{d}$

17.2 Applications of Diffraction, Interference, and Coherence

31. Compare the interference patterns of single-slit diffraction, double-slit diffraction, and a diffraction grating.

a. All three interference pattern produce identical bands.
b. A double slit produces the sharpest and most distinct bands.
c. A single slit produces the sharpest and most distinct bands.
d. The diffraction grating produces the sharpest and most distinct bands.

32. An electric current through hydrogen gas produces several distinct wavelengths of visible light. The light is projected onto a diffraction grating having $10{,}000$ lines per centimeter. What are the wavelengths of the hydrogen spectrum if the light forms first-order maxima at angles of 24.2°, 25.7°, 29.1°, and 41.0°?

a. $\lambda_1 = (10^3\ \text{nm})\sin 24.2^\circ = 410\ \text{nm}$
$\lambda_2 = (10^3\ \text{nm})\sin 25.7^\circ = 434\ \text{nm}$
$\lambda_3 = (10^3\ \text{nm})\sin 29.1^\circ = 486\ \text{nm}$
$\lambda_4 = (10^3\ \text{nm})\sin 41.0^\circ = 656\ \text{nm}$
b. $\lambda_1 = (10^3\ \text{nm})\sin 41.0^\circ = 410\ \text{nm}$
$\lambda_2 = (10^3\ \text{nm})\sin 25.7^\circ = 434\ \text{nm}$
$\lambda_3 = (10^3\ \text{nm})\sin 29.1^\circ = 486\ \text{nm}$
$\lambda_4 = (10^3\ \text{nm})\sin 24.2^\circ = 656\ \text{nm}$
c. $\lambda_1 = (10^3\ \text{nm})\sin 24.2^\circ = 410\ \text{nm}$
$\lambda_2 = (10^3\ \text{nm})\sin 29.1^\circ = 434\ \text{nm}$
$\lambda_3 = (10^3\ \text{nm})\sin 25.7^\circ = 486\ \text{nm}$
$\lambda_4 = (10^3\ \text{nm})\sin 41.0^\circ = 656\ \text{nm}$
d. $\lambda_1 = (10^3\ \text{nm})\sin 41.0^\circ = 410\ \text{nm}$
$\lambda_2 = (10^3\ \text{nm})\sin 29.1^\circ = 434\ \text{nm}$
$\lambda_3 = (10^3\ \text{nm})\sin 25.7^\circ = 486\ \text{nm}$
$\lambda_4 = (10^3\ \text{nm})\sin 24.2^\circ = 656\ \text{nm}$

CHAPTER 18
Static Electricity

Figure 18.1 This child's hair contains an imbalance of electrical charge (commonly called *static electricity*), which causes it to stand on end. The sliding motion stripped electrons away from the child's body, leaving him with an excess of positive charges, which repel each other along each strand of hair. (credit: Ken Bosma, Wikimedia Commons)

Chapter Outline

18.1 Electrical Charges, Conservation of Charge, and Transfer of Charge

18.2 Coulomb's law

18.3 Electric Field

18.4 Electric Potential

18.5 Capacitors and Dielectrics

INTRODUCTION You may have been introduced to static electricity like the child sliding down the slide in the opening photograph (Figure 18.1). The *zap* that he is likely to receive if he touches a playmate or parent tends to bring home the lesson. But static electricity is more than just fun and games—it is put to use in many industries. The forces between electrically charged particles are used in technologies such as printers, pollution filters, and spray guns used for painting cars and trucks. Static *electricity* is the study of phenomena that involve an imbalance of electrical charge. Although creating this imbalance typically requires moving charge around, once the imbalance is created, it often remains static for a long time. The study of charge in motion is called *electromagnetism* and will be covered in a later chapter. What is electrical charge, how is it associated

with objects, and what forces does it create? These are just some of the questions that this chapter addresses.

18.1 Electrical Charges, Conservation of Charge, and Transfer of Charge

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Describe positive and negative electric charges
- Use conservation of charge to calculate quantities of charge transferred between objects
- Characterize materials as conductors or insulators based on their electrical properties
- Describe electric polarization and charging by induction

Section Key Terms

conduction	conductor	electron	induction
insulator	law of conservation of charge	polarization	proton

Electric Charge

You may know someone who has an *electric* personality, which usually means that other people are attracted to this person. This saying is based on electric charge, which is a property of matter that causes objects to attract or repel each other. Electric charge comes in two varieties, which we call *positive* and *negative.* Like charges repel each other, and unlike charges attract each other. Thus, two positive charges repel each other, as do two negative charges. A positive charge and a negative charge attract each other.

How do we know there are two types of electric charge? When various materials are rubbed together in controlled ways, certain combinations of materials always result in a net charge of one type on one material and a net charge of the opposite type on the other material. By convention, we call one type of charge positive and the other type negative. For example, when glass is rubbed with silk, the glass becomes positively charged and the silk negatively charged. Because the glass and silk have opposite charges, they attract one another like clothes that have rubbed together in a dryer. Two glass rods rubbed with silk in this manner will repel one another, because each rod has positive charge on it. Similarly, two silk cloths rubbed in this manner will repel each other, because both cloths have negative charge. Figure 18.2 shows how these simple materials can be used to explore the nature of the force between charges.

Figure 18.2 A glass rod becomes positively charged when rubbed with silk, whereas the silk becomes negatively charged. (a) The glass rod is attracted to the silk, because their charges are opposite. (b) Two similarly charged glass rods repel. (c) Two similarly charged silk cloths repel.

It took scientists a long time to discover what lay behind these two types of charges. The word *electric* itself comes from the Greek word *elektron* for amber, because the ancient Greeks noticed that amber, when rubbed by fur, attracts dry straw. Almost 2,000 years later, the English physicist William Gilbert proposed a model that explained the effect of electric charge as being due to a mysterious electrical fluid that would pass from one object to another. This model was debated for several hundred years, but it was finally put to rest in 1897 by the work of the English physicist J. J. Thomson and French physicist Jean Perrin. Along with many others, Thomson and Perrin were studying the mysterious *cathode rays* that were known at the time to consist of particles smaller than the smallest atom. Perrin showed that cathode rays actually carried negative electrical charge. Later, Thomson's work led him to declare, "I can see no escape from the conclusion that [cathode rays] are charges of negative

electricity carried by particles of matter."

It took several years of further experiments to confirm Thomson's interpretation of the experiments, but science had in fact discovered the particle that carries the fundamental unit of negative electrical charge. We now know this particle as the **electron**.

Atoms, however, were known to be electrically neutral, which means that they carry the same amount of positive and negative charge, so their net charge is zero. Because electrons are negative, some other part of the atom must contain positive charge. Thomson put forth what is called the *plum pudding model*, in which he described atoms as being made of thousands of electrons swimming around in a nebulous mass of positive charge, as shown by the left-side image of Figure 18.3. His student, Ernest Rutherford, originally believed that this model was correct and used it (along with other models) to try to understand the results of his experiments bombarding gold foils with *alpha* particles (i.e., helium atoms stripped of their electrons). The results, however, did not confirm Thomson's model but rather destroyed it! Rutherford found that most of the space occupied by the gold atoms was actually empty and that almost all of the matter of each atom was concentrated into a tiny, extremely dense nucleus, as shown by the right-side image of Figure 18.3. The atomic nucleus was later found to contain particles called **protons**, each of which carries a unit of positive electric charge.[1]

Figure 18.3 The left drawing shows Thompson's plum-pudding model, in which the electrons swim around in a nebulous mass of positive charge. The right drawing shows Rutherford's model, in which the electrons orbit around a tiny, massive nucleus. Note that the size of the nucleus is vastly exaggerated in this drawing. Were it drawn to scale with respect to the size of the electron orbits, the nucleus would not be visible to the naked eye in this drawing. Also, as far as science can currently detect, electrons are point particles, which means that they have no size at all!

Protons and electrons are thus the fundamental particles that carry electric charge. Each proton carries one unit of positive charge, and each electron carries one unit of negative charge. To the best precision that modern technology can provide, the charge carried by a proton is *exactly* the opposite of that carried by an electron. The SI unit for electric charge is the *coulomb* (abbreviated as "C"), which is named after the French physicist Charles Augustin de Coulomb, who studied the force between charged objects. The proton carries $+1.602 \times 10^{-19}\text{C}$. and the electron carries $-1.602 \times 10^{-19}\text{C}$, . The number n of protons required to make +1.00 C is

$$n = 1.00\text{ C} \times \frac{1\text{ proton}}{1.602 \times 10^{-19}\text{C}} = 6.25 \times 10^{18}\text{protons.} \tag{18.1}$$

The same number of electrons is required to make –1.00 C of electric charge. The fundamental unit of charge is often represented as e. Thus, the charge on a proton is e, and the charge on an electron is $-e$. Mathematically, $e = +1.602 \times 10^{-19}\text{C}$.

LINKS TO PHYSICS

Measuring the Fundamental Electric Charge

The American physicist Robert Millikan (1868–1953) and his student Harvey Fletcher (1884–1981) were the first to make a relatively accurate measurement of the fundamental unit of charge on the electron. They designed what is now a classic

1Protons were later found to contain sub particles called *quarks*, which have fractional electric charge. But that is another story that we leave for subsequent physics courses.

experiment performed by students. The Millikan oil-drop experiment is shown in Figure 18.4. The experiment involves some concepts that will be introduced later, but the basic idea is that a fine oil mist is sprayed between two plates that can be charged with a known amount of opposite charge. Some oil drops accumulate some excess negative charge when being sprayed and are attracted to the positive charge of the upper plate and repelled by the negative charge on the lower plate. By tuning the charge on these plates until the weight of the oil drop is balanced by the electric forces, the net charge on the oil drop can be determined quite precisely.

Figure 18.4 The oil-drop experiment involved spraying a fine mist of oil between two metal plates charged with opposite charges. By knowing the mass of the oil droplets and adjusting the electric charge on the plates, the charge on the oil drops can be determined with precision.

Millikan and Fletcher found that the drops would accumulate charge in discrete units of about $-1.59 \times 10^{-19}\,\text{C}$, which is within 1 percent of the modern value of $-1.60 \times 10^{-19}\,\text{C}$. Although this difference may seem quite small, it is actually five times greater than the possible error Millikan reported for his results!

Because the charge on the electron is a fundamental constant of nature, determining its precise value is very important for all of science. This created pressure on Millikan and others after him that reveals some equally important aspects of human nature.

First, Millikan took sole credit for the experiment and was awarded the 1923 Nobel Prize in physics for this work, although his student Harvey Fletcher apparently contributed in significant ways to the work. Just before his death in 1981, Fletcher divulged that Millikan coerced him to give Millikan sole credit for the work, in exchange for which Millikan promoted Fletcher's career at Bell Labs.

Another great scientist, Richard Feynman, points out that many scientists who measured the fundamental charge after Millikan were reluctant to report values that differed much from Millikan's value. History shows that later measurements slowly crept up from Millikan's value until settling on the modern value. Why did they not immediately find the error and correct the value, asks Feynman. Apparently, having found a value higher than the much-respected value found by Millikan, scientists would look for possible mistakes that might lower their value to make it agree better with Millikan's value. This reveals the important psychological weight carried by preconceived notions and shows how hard it is to refute them. Scientists, however devoted to logic and data they may be, are apparently just as vulnerable to this aspect of human nature as everyone else. The lesson here is that, although it is good to be skeptical of new results, you should not discount them just because they do not agree with conventional wisdom. If your reasoning is sound and your data are reliable, the conclusion demanded by the data must be seriously considered, even if that conclusion disagrees with the commonly accepted *truth*.

GRASP CHECK

Suppose that Millikan observed an oil drop carrying three fundamental units of charge. What would be the net charge on this oil drop?

a. -4.81×10^{-19} C
b. -1.602×10^{-19} C
c. 1.602×10^{-19} C
d. 4.81×10^{-19} C

Snap Lab

Like and Unlike Charges

This activity investigates the repulsion and attraction caused by static electrical charge.

- Adhesive tape
- Nonconducting surface, such as a plastic table or chair

Instructions

Procedure for Part (a)

1. Prepare two pieces of tape about 4 cm long. To make a handle, double over about 0.5 cm at one end so that the sticky side sticks together.
2. Attach the pieces of tape side by side onto a nonmetallic surface, such as a tabletop or the seat of a chair, as shown in Figure 18.5(a).
3. Peel off both pieces of tape and hang them downward, holding them by the handles, as shown in Figure 18.5(b). If the tape bends upward and sticks to your hand, try using a shorter piece of tape, or simply shake the tape so that it no longer sticks to your hand.
4. Now slowly bring the two pieces of tape together, as shown in Figure 18.5(c). What happens?

Figure 18.5

Procedure for Part (b)

5. Stick one piece of tape on the nonmetallic surface, and stick the second piece of tape on top of the first piece, as shown in Figure 18.6(a).
6. Slowly peel off the two pieces by pulling on the handle of the bottom piece.
7. Gently stroke your finger along the top of the second piece of tape (i.e., the nonsticky side), as shown in Figure 18.6(b).
8. Peel the two pieces of tape apart by pulling on their handles, as shown in Figure 18.6(c).
9. Slowly bring the two pieces of tape together. What happens?

Figure 18.6

GRASP CHECK

In step 4, why did the two pieces of tape repel each other? In step 9, why did they attract each other?

a. Like charges attract, while unlike charges repel each other.
b. Like charges repel, while unlike charges attract each other.
c. Tapes having positive charge repel, while tapes having negative charge attract each other.
d. Tapes having negative charge repel, while tapes having positive charge attract each other.

Conservation of Charge

Because the fundamental positive and negative units of charge are carried on protons and electrons, we would expect that the total charge cannot change in any system that we define. In other words, although we might be able to move charge around, we cannot create or destroy it. This should be true provided that we do not create or destroy protons or electrons in our system. In the twentieth century, however, scientists learned how to create and destroy electrons and protons, but they found that charge is still conserved. Many experiments and solid theoretical arguments have elevated this idea to the status of a law. The **law of conservation of charge** says that electrical charge cannot be created or destroyed.

The law of conservation of charge is very useful. It tells us that the net charge in a system is the same before and after any interaction within the system. Of course, we must ensure that no external charge enters the system during the interaction and that no internal charge leaves the system. Mathematically, conservation of charge can be expressed as

$$q_{initial} = q_{final}. \tag{18.2}$$

where q_{initial} is the net charge of the system before the interaction, and $q_{\text{final,}}$ is the net charge after the interaction.

WORKED EXAMPLE

What is the missing charge?

Figure 18.7 shows two spheres that initially have +4 C and +8 C of charge. After an interaction (which could simply be that they touch each other), the blue sphere has +10 C of charge, and the red sphere has an unknown quantity of charge. Use the law of conservation of charge to find the final charge on the red sphere.

Strategy

The net initial charge of the system is $q_{\text{initial}} = +4\text{ C} + 8\text{ C} = +12\text{ C}$. The net final charge of the system is $q_{\text{final}} = +10\text{ C} + q_{\text{red}}$, where q_{red} is the final charge on the red sphere. Conservation of charge tells us that $q_{\text{initial}} = q_{\text{final}}$, so we can solve for q_{red}.

Solution

Equating q_{initial} and q_{final} and solving for q_{red} gives

$$\begin{aligned} q_{\text{initial}} &= q_{\text{final}} \\ +12\text{ C} &= +10\text{ C} + q_{\text{red.}} \\ +2\text{ C} &= q_{\text{red.}} \end{aligned} \tag{18.3}$$

The red sphere has +2 C of charge.

Figure 18.7 Two spheres, one blue and one red, initially have +4 C and +8 C of charge, respectively. After the two spheres interact, the blue sphere has a charge of +10 C. The law of conservation of charge allows us to find the final charge q_{red} on the red sphere.

Discussion

Like all conservation laws, conservation of charge is an accounting scheme that helps us keep track of electric charge.

Practice Problems

1. Which equation describes conservation of charge?
   a. $q_{\text{initial}} = q_{\text{final}} = \text{constant}$
   b. $q_{\text{initial}} = q_{\text{final}} = 0$
   c. $q_{\text{initial}} - q_{\text{final}} = 0$

d. $q_{\text{initial}}/q_{\text{final}} = \text{constant}$

2. An isolated system contains two objects with charges q_1 and q_2. If object 1 loses half of its charge, what is the final charge on object 2?
   a. $\frac{q_2}{2}$
   b. $\frac{3q_2}{2}$
   c. $q_2 - \frac{q_1}{2}$
   d. $q_2 + \frac{q_1}{2}$

Conductors and Insulators

Materials can be classified depending on whether they allow charge to move. If charge can easily move through a material, such as metals, then these materials are called **conductors**. This means that charge can be conducted (i.e., move) through the material rather easily. If charge cannot move through a material, such as rubber, then this material is called an **insulator**.

Most materials are insulators. Their atoms and molecules hold on more tightly to their electrons, so it is difficult for electrons to move between atoms. However, it is not impossible. With enough energy, it is possible to force electrons to move through an insulator. However, the insulator is often physically destroyed in the process. In metals, the outer electrons are loosely bound to their atoms, so not much energy is required to make electrons move through metal. Such metals as copper, silver, and aluminum are good conductors. Insulating materials include plastics, glass, ceramics, and wood.

The conductivity of some materials is intermediate between conductors and insulators. These are called *semiconductors*. They can be made conductive under the right conditions, which can involve temperature, the purity of the material, and the force applied to push electrons through them. Because we can control whether semiconductors are conductors or insulators, these materials are used extensively in computer chips. The most commonly used semiconductor is silicon. Figure 18.8 shows various materials arranged according to their ability to conduct electrons.

Figure 18.8 Materials can be arranged according to their ability to conduct electric charge. The slashes on the arrow mean that there is a very large gap in conducting ability between conductors, semiconductors, and insulators, but the drawing is compressed to fit on the page. The numbers below the materials give their *resistivity* in Ω•m (which you will learn about below). The resistivity is a measure of how hard it is to make charge move through a given material.

What happens if an excess negative charge is placed on a conducting object? Because like charges repel each other, they will push against each other until they are as far apart as they can get. Because the charge can move in a conductor, it moves to the outer surfaces of the object. Figure 18.9(a) shows schematically how an excess negative charge spreads itself evenly over the outer surface of a metal sphere.

What happens if the same is done with an insulating object? The electrons still repel each other, but they are not able to move, because the material is an insulator. Thus, the excess charge stays put and does not distribute itself over the object. Figure 18.9(b) shows this situation.

Figure 18.9 (a) A conducting sphere with excess negative charge (i.e., electrons). The electrons repel each other and spread out to cover the outer surface of the sphere. (b) An insulating sphere with excess negative charge. The electrons cannot move, so they remain in their original positions.

Transfer and Separation of Charge

Most objects we deal with are electrically neutral, which means that they have the same amount of positive and negative charge. However, transferring negative charge from one object to another is fairly easy to do. When negative charge is transferred from one object to another, an excess of positive charge is left behind. How do we know that the negative charge is the mobile charge? The positive charge is carried by the proton, which is stuck firmly in the nucleus of atoms, and the atoms are stuck in place in solid materials. Electrons, which carry the negative charge, are much easier to remove from their atoms or molecules and can therefore be transferred more easily.

Electric charge can be transferred in several manners. One of the simplest ways to transfer charge is charging by contact, in which the surfaces of two objects made of different materials are placed in close contact. If one of the materials holds electrons more tightly than the other, then it takes some electrons with it when the materials are separated. Rubbing two surfaces together increases the transfer of electrons, because it creates a closer contact between the materials. It also serves to present *fresh* material with a full supply of electrons to the other material. Thus, when you walk across a carpet on a dry day, your shoes rub against the carpet, and some electrons are removed from the carpet by your shoes. The result is that you have an excess of negative charge on your shoes. When you then touch a doorknob, some of your excess of electrons transfer to the neutral doorknob, creating a small spark.

Touching the doorknob with your hand demonstrates a second way to transfer electric charge, which is charging by **conduction**. This transfer happens because like charges repel, and so the excess electrons that you picked up from the carpet want to be as far away from each other as possible. Some of them move to the doorknob, where they will distribute themselves over the outer surface of the metal. Another example of charging by conduction is shown in the top row of Figure 18.10. A metal sphere with 100 excess electrons touches a metal sphere with 50 excess electrons, so 25 electrons from the first sphere transfer to the second sphere. Each sphere finishes with 75 excess electrons.

The same reasoning applies to the transfer of positive charge. However, because positive charge essentially cannot move in solids, it is transferred by moving negative charge in the opposite direction. For example, consider the bottom row of Figure 18.10. The first metal sphere has 100 excess protons and touches a metal sphere with 50 excess protons, so the second sphere transfers 25 electrons to the first sphere. These 25 extra electrons will electrically cancel 25 protons so that the first metal sphere is left with 75 excess protons. This is shown in the bottom row of Figure 18.10. The second metal sphere lost 25 electrons so it has 25 more excess protons, for a total of 75 excess protons. The end result is the same if we consider that the first ball transferred a net positive charge equal to that of 25 protons to the first ball.

Figure 18.10 In the top row, a metal sphere with 100 excess electrons transfers 25 electrons to a metal sphere with an excess of 50 electrons. After the transfer, both spheres have 75 excess electrons. In the bottom row, a metal sphere with 100 excess protons receives 25 electrons from a ball with 50 excess protons. After the transfer, both spheres have 75 excess protons.

In this discussion, you may wonder how the excess electrons originally got from your shoes to your hand to create the spark when you touched the doorknob. The answer is that *no* electrons actually traveled from your shoes to your hands. Instead, because like charges repel each other, the excess electrons on your shoe simply pushed away some of the electrons in your feet. The electrons thus dislodged from your feet moved up into your leg and in turn pushed away some electrons in your leg. This process continued through your whole body until a distribution of excess electrons covered the extremities of your body. Thus your head, your hands, the tip of your nose, and so forth all received their doses of excess electrons that had been pushed out of their normal positions. All this was the result of electrons being pushed out of your feet by the excess electrons on your shoes.

This type of charge separation is called **polarization**. As soon as the excess electrons leave your shoes (by rubbing off onto the floor or being carried away in humid air), the distribution of electrons in your body returns to normal. Every part of your body is again electrically neutral (i.e., zero excess charge).

The phenomenon of polarization is seen in . The child has accumulated excess positive charge by sliding on the slide. This excess charge repels itself and so becomes distributed over the extremities of the child's body, notably in his hair. As a result, the hair stands on end, because the excess negative charge on each strand repels the excess positive charge on neighboring strands.

Polarization can be used to charge objects. Consider the two metallic spheres shown in Figure 18.11. The spheres are electrically neutral, so they carry the same amounts of positive and negative charge. In the top picture (Figure 18.11(a)), the two spheres are touching, and the positive and negative charge is evenly distributed over the two spheres. We then approach a glass rod that carries an excess positive charge, which can be done by rubbing the glass rod with silk, as shown in Figure 18.11(b). Because opposite charges attract each other, the negative charge is attracted to the glass rod, leaving an excess positive charge on the opposite side of the right sphere. This is an example of charging by **induction**, whereby a charge is created by approaching a charged object with a second object to create an unbalanced charge in the second object. If we then separate the two spheres, as shown in Figure 18.11(c), the excess charge is stuck on each sphere. The left sphere now has an excess negative charge, and the right sphere has an excess positive charge. Finally, in the bottom picture, the rod is removed, and the opposite charges attract each other, so they move as close together as they can get.

Figure 18.11 (a) Two neutral conducting spheres are touching each other, so the charge is evenly spread over both spheres. (b) A positively charged rod approaches, which attracts negative charges, leaving excess positive charge on the right sphere. (c) The spheres are separated. Each sphere now carries an equal magnitude of excess charge. (d) When the positively charged rod is removed, the excess negative charge on the left sphere is attracted to the excess positive charge on the right sphere.

FUN IN PHYSICS

Create a Spark in a Science Fair

Van de Graaff generators are devices that are used not only for serious physics research but also for demonstrating the physics of static electricity at science fairs and in classrooms. Because they deliver relatively little electric current, they can be made safe for use in such environments. The first such generator was built by Robert Van de Graaff in 1931 for use in nuclear physics research. Figure 18.12 shows a simplified sketch of a Van de Graaff generator.

Van de Graaff generators use smooth and pointed surfaces and conductors and insulators to generate large static charges. In the version shown in Figure 18.12, electrons are "sprayed" from the tips of the lower comb onto a moving belt, which is made of an insulating material like, such as rubber. This technique of charging the belt is akin to charging your shoes with electrons by walking across a carpet. The belt raises the charges up to the upper comb, where they transfer again, akin to your touching the doorknob and transferring your charge to it. Because like charges repel, the excess electrons all rush to the outer surface of the globe, which is made of metal (a conductor). Thus, the comb itself never accumulates too much charge, because any charge it gains is quickly depleted by the charge moving to the outer surface of the globe.

Figure 18.12 Van de Graaff generators transfer electrons onto a metallic sphere, where the electrons distribute themselves uniformly over the outer surface.

Van de Graaff generators are used to demonstrate many interesting effects caused by static electricity. By touching the globe, a person gains excess charge, so his or her hair stands on end, as shown in Figure 18.13. You can also create mini lightning bolts by moving a neutral conductor toward the globe. Another favorite is to pile up aluminum muffin tins on top of the uncharged globe, then turn on the generator. Being made of conducting material, the tins accumulate excess charge. They then repel each other and fly off the globe one by one. A quick Internet search will show many examples of what you can do with a Van de Graaff generator.

Figure 18.13 The man touching the Van de Graaff generator has excess charge, which spreads over his hair and repels hair strands from his neighbors. (credit: Jon "ShakataGaNai" Davis)

GRASP CHECK

Why don't the electrons stay on the rubber belt when they reach the upper comb?

a. The upper comb has no excess electrons, and the excess electrons in the rubber belt get transferred to the comb by contact.
b. The upper comb has no excess electrons, and the excess electrons in the rubber belt get transferred to the comb by conduction.
c. The upper comb has excess electrons, and the excess electrons in the rubber belt get transferred to the comb by conduction.
d. The upper comb has excess electrons, and the excess electrons in the rubber belt get transferred to the comb by contact.

Virtual Physics

Balloons and Static Electricity

Click to view content (http://www.openstax.org/l/28balloons)

This simulation allows you to observe negative charge accumulating on a balloon as you rub it against a sweater. You can then observe how two charged balloons interact and how they cause polarization in a wall.

GRASP CHECK

Click the reset button, and start with two balloons. Charge a first balloon by rubbing it on the sweater, and then move it toward the second balloon. Why does the second balloon not move?

a. The second balloon has an equal number of positive and negative charges.
b. The second balloon has more positive charges than negative charges.
c. The second balloon has more negative charges than positive charges.
d. The second balloon is positively charged and has polarization.

Snap Lab

Polarizing Tap Water

This lab will demonstrate how water molecules can easily be polarized.

- Plastic object of small dimensions, such as comb or plastic stirrer
- Source of tap water

Instructions

Procedure

1. Thoroughly rub the plastic object with a dry cloth.
2. Open the faucet just enough to let a smooth filament of water run from the tap.
3. Move an edge of the charged plastic object toward the filament of running water.

What do you observe? What happens when the plastic object touches the water filament? Can you explain your observations?

GRASP CHECK

Why does the water curve around the charged object?

a. The charged object induces uniform positive charge on the water molecules.
b. The charged object induces uniform negative charge on the water molecules.
c. The charged object attracts the polarized water molecules and ions that are dissolved in the water.
d. The charged object depolarizes the water molecules and the ions dissolved in the water.

WORKED EXAMPLE

Charging Ink Droplets

Electrically neutral ink droplets in an ink-jet printer pass through an electron beam created by an electron gun, as shown in Figure 18.14. Some electrons are captured by the ink droplet, so that it becomes charged. After passing through the electron beam, the net charge of the ink droplet is $q_{\text{inkdrop}} = -1 \times 10^{-10}\text{C}$. How many electrons are captured by the ink droplet?

Figure 18.14 Electrons from an *electron gun* charge a passing ink droplet.

STRATEGY

A single electron carries a charge of $q_{e^-} = -1.602 \times 10^{-19}\,\text{C}$. Dividing the net charge of the ink droplet by the charge q_{e^-} of a single electron will give the number of electrons captured by the ink droplet.

Solution

The number n of electrons captured by the ink droplet are

$$n = \frac{q_{\text{inkdrop}}}{q_{e^-}} = \frac{-1 \times 10^{-10}\,\text{C}}{-1.602 \times 10^{-19}\,\text{C}} = 6 \times 108. \tag{18.4}$$

Discussion

This is almost a billion electrons! It seems like a lot, but it is quite small compared to the number of atoms in an ink droplet, which number about 10^{16}. Thus, each extra electron is shared between about $10^{16}/\left(6 \times 10^{8}\right) \approx 10^{7}$ atoms.

Practice Problems

3. How many protons are needed to make 1 nC of charge? 1 nC = 10–9 C
   a. 1.6×10^{-28}
   b. 1.6×10^{-10}
   c. 3×10^{9}
   d. 6×10^{9}

4. In a physics lab, you charge up three metal spheres, two with $+3\,\text{nC}$ and one with $-5\,\text{nC}$. When you bring all three spheres together so that they all touch one another, what is the total charge on the three spheres?
   a. $+1\,\text{nC}$
   b. $+3\,\text{nC}$
   c. $+5\,\text{nC}$
   d. $+6\,\text{nC}$

Check Your Understanding

5. How many types of electric charge exist?
   a. one type
   b. two types
   c. three types
   d. four types

6. Which are the two main electrical classifications of materials based on how easily charges can move through them?

a. conductor and insulator
b. semiconductor and insulator
c. conductor and superconductor
d. conductor and semiconductor

7. True or false—A polarized material must have a nonzero net electric charge.
a. true
b. false

8. Describe the force between two positive point charges that interact.
a. The force is attractive and acts along the line joining the two point charges.
b. The force is attractive and acts tangential to the line joining the two point charges.
c. The force is repulsive and acts along the line joining the two point charges.
d. The force is repulsive and acts tangential to the line joining the two point charges.

9. How does a conductor differ from an insulator?
a. Electric charges move easily in an insulator but not in a conducting material.
b. Electric charges move easily in a conductor but not in an insulator.
c. A conductor has a large number of electrons.
d. More charges are in an insulator than in a conductor.

10. True or false—Charging an object by polarization requires touching it with an object carrying excess charge.
a. true
b. false

18.2 Coulomb's law

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Describe Coulomb's law verbally and mathematically
- Solve problems involving Coulomb's law

Section Key Terms

Coulomb's law	inverse-square law

More than 100 years before Thomson and Rutherford discovered the fundamental particles that carry positive and negative electric charges, the French scientist Charles-Augustin de Coulomb mathematically described the force between charged objects. Doing so required careful measurements of forces between charged spheres, for which he built an ingenious device called a *torsion balance*.

This device, shown in Figure 18.15, contains an insulating rod that is hanging by a thread inside a glass-walled enclosure. At one end of the rod is the metallic sphere A. When no charge is on this sphere, it touches sphere B. Coulomb would touch the spheres with a third metallic ball (shown at the bottom of the diagram) that was charged. An unknown amount of charge would distribute evenly between spheres A and B, which would then repel each other, because like charges repel. This force would cause sphere A to rotate away from sphere B, thus twisting the wire until the torsion in the wire balanced the electrical force. Coulomb then turned the knob at the top, which allowed him to rotate the thread, thus bringing sphere A closer to sphere B. He found that bringing sphere A twice as close to sphere B required increasing the torsion by a factor of four. Bringing the sphere three times closer required a ninefold increase in the torsion. From this type of measurement, he deduced that the electrical force between the spheres was inversely proportional to the distance squared between the spheres. In other words,

$$F \propto \frac{1}{r^2}, \tag{18.5}$$

where r is the distance between the spheres.

An electrical charge distributes itself equally between two conducting spheres of the same size. Knowing this allowed Coulomb to divide an unknown charge in half. Repeating this process would produce a sphere with one quarter of the initial charge, and

so on. Using this technique, he measured the force between spheres A and B when they were charged with different amounts of charge. These measurements led him to deduce that the force was proportional to the charge on each sphere, or

$$F \propto q_A q_B, \tag{18.6}$$

where q_A is the charge on sphere A, and q_B is the charge on sphere B.

Figure 18.15 A drawing of Coulomb's torsion balance, which he used to measure the electrical force between charged spheres. (credit: Charles-Augustin de Coulomb)

Combining these two proportionalities, he proposed the following expression to describe the force between the charged spheres.

$$F = \frac{kq_1 q_2}{r^2} \tag{18.7}$$

This equation is known as **Coulomb's law**, and it describes the electrostatic force between charged objects. The constant of proportionality k is called *Coulomb's constant*. In SI units, the constant k has the value $k = 8.99 \times 10^9 \text{N} \cdot \text{m}^2/\text{C}^2$.

The direction of the force is along the line joining the centers of the two objects. If the two charges are of opposite signs, Coulomb's law gives a negative result. This means that the force between the particles is attractive. If the two charges have the same signs, Coulomb's law gives a positive result. This means that the force between the particles is repulsive. For example, if both q_1 and q_2 are negative or if both are positive, the force between them is repulsive. This is shown in Figure 18.16(a). If q_1 is a negative charge and q_2 is a positive charge (or vice versa), then the charges are different, so the force between them is attractive. This is shown in Figure 18.16(b).

Figure 18.16 The magnitude of the electrostatic force F between point charges q_1 and q_2 separated by a distance r is given by Coulomb's law. Note that Newton's third law (every force exerted creates an equal and opposite force) applies as usual—the force ($F_{1,2}$) on q_1 is equal in magnitude and opposite in direction to the force ($F_{2,1}$) it exerts on q_2. (a) Like charges. (b) Unlike charges.

Note that Coulomb's law applies only to charged objects that are not moving with respect to each other. The law says that the force is proportional to the amount of charge on each object and inversely proportional to the square of the distance between the objects. If we double the charge q_1, for instance, then the force is doubled. If we double the distance between the objects, then the force between them *decreases* by a factor of $2^2 = 4$. Although Coulomb's law is true in general, it is easiest to apply to spherical objects or to objects that are much smaller than the distance between the objects (in which case, the objects can be approximated as spheres).

Coulomb's law is an example of an **inverse-square law**, which means the force depends on the square of the denominator. Another inverse-square law is Newton's law of universal gravitation, which is $F = Gm_1m_2/r^2$. Although these laws are similar, they differ in two important respects: (i) The gravitational constant G is much, much smaller than k ($G = 6.67 \times 10^{-11}\,\mathrm{m^3/kg \cdot s^2}$); and (ii) only one type of mass exists, whereas two types of electric charge exist. These two differences explain why gravity is so much weaker than the electrostatic force and why gravity is only attractive, whereas the electrostatic force can be attractive or repulsive.

Finally, note that Coulomb measured the distance between the spheres from the centers of each sphere. He did not explain this assumption in his original papers, but it turns out to be valid. From *outside* a uniform spherical distribution of charge, it can be treated as if all the charge were located at the center of the sphere.

WATCH PHYSICS

Electrostatics (part 1): Introduction to charge and Coulomb's law

This video explains the basics of Coulomb's law. Note that the lecturer uses d for the distance between the center of the particles instead of r.

Click to view content (https://www.openstax.org/l/28coulomb)

GRASP CHECK

True or false—If one particle carries a positive charge and another carries a negative charge, then the force between them is attractive.

a. true
b. false

Snap Lab

Hovering plastic

In this lab, you will use electrostatics to hover a thin piece of plastic in the air.

- Balloon
- Light plastic bag (e.g., produce bag from grocery store)

Instructions

Procedure

1. Cut the plastic bag to make a plastic loop about 2 inches wide.
2. Inflate the balloon.
3. Charge the balloon by rubbing it on your clothes.
4. Charge the plastic loop by placing it on a nonmetallic surface and rubbing it with a cloth.
5. Hold the balloon in one hand, and in the other hand hold the plastic loop above the balloon. If the loop clings too much to your hand, recruit a friend to hold the strip above the balloon with both hands. Now let go of the plastic loop, and maneuver the balloon under the plastic loop to keep it hovering in the air above the balloon.

GRASP CHECK

How does the balloon keep the plastic loop hovering?

a. The balloon and the loop are both negatively charged. This will help the balloon keep the plastic loop hovering.
b. The balloon is charged, while the plastic loop is neutral.This will help the balloon keep the plastic loop hovering.
c. The balloon and the loop are both positively charged. This will help the balloon keep the plastic loop hovering.
d. The balloon is positively charged, while the plastic loop is negatively charged. This will help the balloon keep the plastic loop hovering.

WORKED EXAMPLE

Using Coulomb's law to find the force between charged objects

Suppose Coulomb measures a force of $20 \times 10^{-6}\,\text{N}$ between the two charged spheres when they are separated by 5.0 cm. By turning the dial at the top of the torsion balance, he approaches the spheres so that they are separated by 3.0 cm. Which force does he measure now?

STRATEGY

Apply Coulomb's law to the situation before and after the spheres are brought closer together. Although we do not know the charges on the spheres, we do know that they remain the same. We call these unknown but constant charges q_1 and q_2. Because these charges appear as a product in Coulomb's law, they form a single unknown. We thus have two equations and two unknowns, which we can solve. The first unknown is the force (which we call F_f) when the spheres are 3.0 cm apart, and the second is $q_1 q_2$.

Use the following notation: When the charges are 5.0 cm apart, the force is $F_\text{i} = 20 \times 10^{-6}\,\text{N}$ and $r_\text{i} = 5.0\text{ cm} = 0.050\text{ m}$, where the subscript *i* means *initial*. Once the charges are brought closer together, we know $r_\text{f} = 3.0\text{ cm} = 0.030\text{ m}$, where the subscript *f* means *final*.

Solution

Coulomb's law applied to the spheres in their initial positions gives

$$F_\text{i} = \frac{kq_1 q_2}{r_\text{i}^2}. \tag{18.8}$$

Coulomb's law applied to the spheres in their final positions gives

$$F_\text{f} = \frac{kq_1 q_2}{r_\text{f}^2}. \tag{18.9}$$

Dividing the second equation by the first and solving for the final force F_f leads to

$$\begin{aligned} \frac{F_\text{f}}{F_\text{i}} &= \frac{kq_1 q_2/r_\text{f}^2}{kq_1 q_2/r_\text{i}^2}. \\ &= \frac{r_\text{i}^2}{r_\text{f}^2} \\ F_\text{f} &= F_\text{i}\frac{r_\text{i}^2}{r_\text{f}^2}. \end{aligned} \tag{18.10}$$

Inserting the known quantities yields

$$\begin{aligned} F_\text{f} &= F_\text{i}\frac{r_\text{i}^2}{r_\text{f}^2} \\ &= \left(20 \times 10^{-6}\,\text{N}\right)\frac{(0.050\text{ m})^2}{(0.030\text{ m})^2} \\ &= 56 \times 10^{-6}\,\text{N} \\ &= 5.6 \times 10^{-5}\,\text{N}. \end{aligned} \tag{18.11}$$

The force acts along the line joining the centers of the spheres. Because the same type of charge is on each sphere, the force is repulsive.

Discussion

As expected, the force between the charges is greater when they are 3.0 cm apart than when they are 5.0 cm apart. Note that although it is a good habit to convert cm to m (because the constant *k* is in SI units), it is not necessary in this problem, because the distances cancel out.

We can also solve for the second unknown $|q_1 q_2|$. By using the first equation, we find

$$\begin{aligned} F_{\text{f}} &= \frac{kq_1q_2}{r_{\text{i}}^2} \\ q_1q_2 &= \frac{F_{\text{i}}r_{\text{i}}^2}{k} \\ &= \frac{(20 \times 10^{-6}\text{N})(0.050\ \text{m})^2}{8.99 \times 10^9\text{N}\cdot\text{m}^2/\text{C}^2} \\ &= 5.6 \times 10^{-18}\text{C}^2 \end{aligned} \tag{18.12}$$

Note how the units cancel in the second-to-last line. Had we not converted cm to m, this would not occur, and the result would be incorrect. Finally, because the charge on each sphere is the same, we can further deduce that

$$q_1 = q_2 = \pm\sqrt{5.6 \times 10^{-18}\text{C}^2} = \pm 2.4\ \text{nC}. \tag{18.13}$$

WORKED EXAMPLE

Using Coulomb's law to find the distance between charged objects

An engineer measures the force between two ink drops by measuring their acceleration and their diameter. She finds that each member of a pair of ink drops exerts a repulsive force of $F = 5.5\ \text{mN}$ on its partner. If each ink drop carries a charge $q_{\text{inkdrop}} = -1 \times 10^{-10}\text{C}$, how far apart are the ink drops?

STRATEGY

We know the force and the charge on each ink drop, so we can solve Coulomb's law for the distance *r* between the ink drops. Do not forget to convert the force into SI units: $F = 5.5\ \text{mN} = 5.5 \times 10^{-3}\text{N}$.

Solution

The charges in Coulomb's law are $q_1 = q_2 = q_{\text{inkdrop}}$, so the numerator in Coulomb's law takes the form $q_1q_2 = q_{\text{inkdrop}}^2$. Inserting this into Coulomb's law and solving for the distance *r* gives

$$\begin{aligned} F &= \frac{kq_{\text{inkdrop}}^2}{r^2} \\ r &= \pm\sqrt{\frac{kq_{\text{inkdrop}}^2}{F}} \\ &= \pm\sqrt{\frac{(8.99 \times 10^9\text{N}\cdot\text{m}^2/\text{C}^2)(-1 \times 10^{-10}\text{C})^2}{5.5\times10^{-3}\text{N}}} \\ &= \pm 1.3 \times 10^{-4}\text{m} \end{aligned} \tag{18.14}$$

or 130 microns (about one-tenth of a millimeter).

Discussion

The plus-minus sign means that we do not know which ink drop is to the right and which is to the left, but that is not important, because both ink drops are the same.

Practice Problems

11. A charge of -4×10^{-9} C is a distance of 3 cm from a charge of 3×10^{-9} C . What is the magnitude and direction of the force between them?

a. 1.2×10^{-4} N, and the force is attractive
b. 1.2×10^{14} N, and the force is attractive
c. 6.74×10^{23} N, and the force is attractive
d. $-\hat{y}$, and the force is attractive

12. Two charges are repelled by a force of 2.0 N. If the distance between them triples, what is the force between the charges?

a. 0.22 N
b. 0.67 N
c. 2.0 N
d. 18.0 N

Check Your Understanding

13. How are electrostatic force and charge related?

a. The force is proportional to the product of two charges.
b. The force is inversely proportional to the product of two charges.
c. The force is proportional to any one of the charges between which the force is acting.
d. The force is inversely proportional to any one of the charges between which the force is acting.

14. Why is Coulomb's law called an inverse-square law?

a. because the force is proportional to the inverse of the distance squared between charges
b. because the force is proportional to the product of two charges
c. because the force is proportional to the inverse of the product of two charges
d. because the force is proportional to the distance squared between charges

18.3 Electric Field

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Calculate the strength of an electric field
- Create and interpret drawings of electric fields

Section Key Terms

electric field	test charge

You may have heard of a *force field* in science fiction movies, where such fields apply forces at particular positions in space to keep a villain trapped or to protect a spaceship from enemy fire. The concept of a *field* is very useful in physics, although it differs somewhat from what you see in movies.

A *field* is a way of conceptualizing and mapping the force that surrounds any object and acts on another object at a distance without apparent physical connection. For example, the gravitational field surrounding Earth and all other masses represents the gravitational force that would be experienced if another mass were placed at a given point within the field. Michael Faraday, an English physicist of the nineteenth century, proposed the concept of an **electric field**. If you know the electric field, then you can easily calculate the force (magnitude and direction) applied to any electric charge that you place in the field.

An electric field is generated by electric charge and tells us the force per unit charge at all locations in space around a charge distribution. The charge distribution could be a single point charge; a distribution of charge over, say, a flat plate; or a more complex distribution of charge. The electric field extends into space around the charge distribution. Now consider placing a test charge in the field. A **test charge** is a positive electric charge whose charge is so small that it does not significantly disturb the charges that create the electric field. The electric field exerts a force on the test charge in a given direction. The force exerted is proportional to the charge of the test charge. For example, if we double the charge of the test charge, the force exerted on it doubles. Mathematically, saying that electric field is the force per unit charge is written as

$$\vec{E} = \frac{\vec{F}}{q_{\text{test}}} \tag{18.15}$$

where we are considering only electric forces. Note that the electric field is a vector field that points in the same direction as the force on the positive test charge. The units of electric field are N/C.

If the electric field is created by a point charge or a sphere of uniform charge, then the magnitude of the force between this point charge Q and the test charge is given by Coulomb's law

$$F = \frac{k\,|Qq_{\text{test}}|}{r^2}$$

where the absolute value is used, because we only consider the magnitude of the force. The magnitude of the electric field is then

$$E = \frac{F}{q_{\text{test}}} = \frac{k\,|Q|}{r^2}. \tag{18.16}$$

This equation gives the magnitude of the electric field created by a point charge Q. The distance r in the denominator is the distance from the point charge, Q, or from the center of a spherical charge, to the point of interest.

If the test charge is removed from the electric field, the electric field still exists. To create a three-dimensional map of the electric field, imagine placing the test charge in various locations in the field. At each location, measure the force on the charge, and use the vector equation $\vec{E} = \vec{F}/q_{\text{test}}$ to calculate the electric field. Draw an arrow at each point where you place the test charge to represent the strength and the direction of the electric field. The length of the arrows should be proportional to the strength of the electric field. If you join together these arrows, you obtain lines. Figure 18.17 shows an image of the three-dimensional electric field created by a positive charge.

Figure 18.17 Three-dimensional representation of the electric field generated by a positive charge.

Just drawing the electric field lines in a plane that slices through the charge gives the two-dimensional electric-field maps shown in Figure 18.18. On the left is the electric field created by a positive charge, and on the right is the electric field created by a negative charge.

Notice that the electric field lines point away from the positive charge and toward the negative charge. Thus, a positive test charge placed in the electric field of the positive charge will be repelled. This is consistent with Coulomb's law, which says that like charges repel each other. If we place the positive charge in the electric field of the negative charge, the positive charge is attracted to the negative charge. The opposite is true for negative test charges. Thus, the direction of the electric field lines is consistent with what we find by using Coulomb's law.

The equation $E = k\,|Q|/r^2$ says that the electric field gets stronger as we approach the charge that generates it. For example, at 2 cm from the charge Q ($r = 2$ cm), the electric field is four times stronger than at 4 cm from the charge ($r = 4$ cm). Looking at Figure 18.17 and Figure 18.18 again, we see that the electric field lines become denser as we approach the charge that generates it. In fact, the density of the electric field lines is proportional to the strength of the electric field!

Figure 18.18 Electric field lines from two point charges. The red point on the left carries a charge of +1 nC, and the blue point on the right carries a charge of −1 nC. The arrows point in the direction that a positive test charge would move. The field lines are denser as you approach the point charge.

Electric-field maps can be made for several charges or for more complicated charge distributions. The electric field due to multiple charges may be found by adding together the electric field from each individual charge. Because this sum can only be a single number, we know that only a single electric-field line can go through any given point. In other words, electric-field lines *cannot* cross each other.

Figure 18.19(a) shows a two-dimensional map of the electric field generated by a charge of $+q$ and a nearby charge of $-q$. The three-dimensional version of this map is obtained by rotating this map about the axis that goes through both charges. A positive

test charge placed in this field would experience a force in the direction of the field lines at its location. It would thus be repelled from the positive charge and attracted to the negative charge. Figure 18.19(b) shows the electric field generated by two charges of $-q$. Note how the field lines tend to repel each other and do not overlap. A positive test charge placed in this field would be attracted to both charges. If you are far from these two charges, where far means much farther than the distance between the charges, the electric field looks like the electric field from a single charge of $-2q$.

Figure 18.19 (a) The electric field generated by a positive point charge (left) and a negative point charge of the same magnitude (right). (b) The electric field generated by two equal negative charges.

Virtual Physics

Probing an Electric Field

Click to view content (http://www.openstax.org/l/28charge-field)

This simulation shows you the electric field due to charges that you place on the screen. Start by clicking the top checkbox in the options panel on the right-hand side to show the electric field. Drag charges from the buckets onto the screen, move them around, and observe the electric field that they form. To see more precisely the magnitude and direction of the electric field, drag an electric-field sensor, or *E-field* sensor from the bottom bucket, and move it around the screen.

GRASP CHECK

Two positive charges are placed on a screen. Which statement describes the electric field produced by the charges?

a. It is constant everywhere.
b. It is zero near each charge.
c. It is zero halfway between the charges.
d. It is strongest halfway between the charges.

WATCH PHYSICS

Electrostatics (part 2): Interpreting electric field

This video explains how to calculate the electric field of a point charge and how to interpret electric-field maps in general. Note

that the lecturer uses *d* for the distance between particles instead of *r*. Note that the point charges are infinitesimally small, so all their charges are focused at a point. When larger charged objects are considered, the distance between the objects must be measured between the center of the objects.

Click to view content (https://www.youtube.com/embed/oYOGrTNgGhE)

GRASP CHECK

True or false—If a point charge has electric field lines that point into it, the charge must be ositive.

a. true
b. false

WORKED EXAMPLE

What is the charge?

Look at the drawing of the electric field in Figure 18.20. What is the relative strength and sign of the three charges?

Figure 18.20 Map of electric field due to three charged particles.

STRATEGY

We know the electric field extends out from positive charge and terminates on negative charge. We also know that the number of electric field lines that touch a charge is proportional to the charge. Charge 1 has 12 fields coming out of it. Charge 2 has six field lines going into it. Charge 3 has 12 field lines going into it.

Solution

The electric-field lines come out of charge 1, so it is a positive charge. The electric-field lines go into charges 2 and 3, so they are negative charges. The ratio of the charges is $q_1 : q_2 : q_3 = +12 : -6 : -12$. Thus, magnitude of charges 1 and 3 is twice that of charge 2.

Discussion

Although we cannot determine the precise charge on each particle, we can get a lot of information from the electric field regarding the magnitude and sign of the charges and where the force on a test charge would be greatest (or least).

WORKED EXAMPLE

Electric field from doorknob

A doorknob, which can be taken to be a spherical metal conductor, acquires a static electricity charge of $q = -1.5\ \text{nC}$. What is the electric field 1.0 cm in front of the doorknob? The diameter of the doorknob is 5.0 cm.

STRATEGY

Because the doorknob is a conductor, the entire charge is distributed on the outside surface of the metal. In addition, because the doorknob is assumed to be perfectly spherical, the charge on the surface is uniformly distributed, so we can treat the doorknob as if all the charge were located at the center of the doorknob. The validity of this simplification will be proved in a later physics course. Now sketch the doorknob, and define your coordinate system. Use $+x$ to indicate the outward direction

perpendicular to the door, with $x = 0$ at the center of the doorknob (as shown in the figure below).

If the diameter of the doorknob is 5.0 cm, its radius is 2.5 cm. We want to know the electric field 1.0 cm from the surface of the doorknob, which is a distance $r = 2.5\text{ cm} + 1.0\text{ cm} = 3.5\text{ cm}$ from the center of the doorknob. We can use the equation $E = \frac{k|Q|}{r^2}$ to find the magnitude of the electric field. The direction of the electric field is determined by the sign of the charge, which is negative in this case.

Solution

Inserting the charge $Q = -1.5\text{ nC} = -1.5 \times 10^{-9}\text{C}$ and the distance $r = 3.5\text{ cm} = 0.035\text{ m}$ into the equation $E = \frac{k|Q|}{r^2}$ gives

$$\begin{aligned} E &= \frac{k|Q|}{r^2} \\ &= \frac{(8.99 \times 10^9\text{N}\cdot\text{m}^2/\text{C}^2)|-1.5 \times 10^{-9}\text{C}|}{(0.035\text{ m})^2} \\ &= 1.1 \times 10^4\text{N/C}. \end{aligned} \tag{18.17}$$

Because the charge is negative, the electric-field lines point toward the center of the doorknob. Thus, the electric field at $x = 3.5\text{ cm}$ is $(-1.1 \times 10^4\text{N/C})\hat{x}$.

Discussion

This seems like an enormous electric field. Luckily, it takes an electric field roughly 100 times stronger ($3 \times 10^6\text{N/C}$) to cause air to break down and conduct electricity. Also, the weight of an adult is about $70\text{ kg} \times 9.8\text{m/s}^2 \approx 700\text{ N}$, so why don't you feel a force on the protons in your hand as you reach for the doorknob? The reason is that your hand contains an equal amount of negative charge, which repels the negative charge in the doorknob. A very small force might develop from polarization in your hand, but you would never notice it.

Practice Problems

15. What is the magnitude of the electric field from 20 cm from a point charge of $q = 33$ nC?

a. 7.4×10^3 N/C
b. 1.48×10^3 N / C
c. 7.4×10^{12} N / C
d. 0

16. A –10 nC charge is at the origin. In which direction does the electric field from the charge point at $x + 10$ cm ?

a. The electric field points away from negative charges.
b. The electric field points toward negative charges.
c. The electric field points toward positive charges.
d. The electric field points away from positive charges.

Check Your Understanding

17. When electric field lines get closer together, what does that tell you about the electric field?

a. The electric field is inversely proportional to the density of electric field lines.
b. The electric field is directly proportional to the density of electric field lines.
c. The electric field is not related to the density of electric field lines.
d. The electric field is inversely proportional to the square root of density of electric field lines.

18. If five electric-field lines come out of a +5 nC charge, how many electric-field lines should come out of a +20 nC charge?
a. five field lines
b. 10 field lines
c. 15 field lines
d. 20 field lines

18.4 Electric Potential

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Explain the similarities and differences between electric potential energy and gravitational potential energy
- Calculate the electric potential difference between two point charges and in a uniform electric field

Section Key Terms

electric potential	electric potential energy

As you learned in studying gravity, a mass in a gravitational field has potential energy, which means it has the potential to accelerate and thereby increase its kinetic energy. This kinetic energy can be used to do work. For example, imagine you want to use a stone to pound a nail into a piece of wood. You first lift the stone high above the nail, which increases the potential energy of the stone-Earth system—because Earth is so large, it does not move, so we usually shorten this by saying simply that the potential energy of the stone increases. When you drop the stone, gravity converts the potential energy into kinetic energy. When the stone hits the nail, it does work by pounding the nail into the wood. The gravitational potential energy is the work that a mass can potentially do by virtue of its position in a gravitational field. Potential energy is a very useful concept, because it can be used with conservation of energy to calculate the motion of masses in a gravitational field.

Electric potential energy works much the same way, but it is based on the electric field instead of the gravitational field. By virtue of its position in an electric field, a charge has an electric potential energy. If the charge is free to move, the force due to the electric field causes it to accelerate, so its potential energy is converted to kinetic energy, just like a mass that falls in a gravitational field. This kinetic energy can be used to do work. The electric potential energy is the work that a charge can do by virtue of its position in an electric field.

The analogy between gravitational potential energy and electric potential energy is depicted in Figure 18.21. On the left, the ball-Earth system gains gravitational potential energy when the ball is higher in Earth's gravitational field. On the right, the two-charge system gains electric potential energy when the positive charge is farther from the negative charge.

Figure 18.21 On the left, the gravitational field points toward Earth. The higher the ball is in the gravitational field, the higher the potential energy is of the Earth-ball system. On the right, the electric field points toward a negative charge. The farther the positive charge is from the negative charge, the higher the potential energy is of the two-charge system.

Let's use the symbol U_G to denote gravitational potential energy. When a mass falls in a gravitational field, its gravitational potential energy decreases. Conservation of energy tells us that the work done by the gravitational field to make the mass accelerate must equal the loss of potential energy of the mass. If we use the symbol $W_{\text{donebygravity}}$ to denote this work, then

$$-\Delta U_{\text{G}} = W_{\text{donebygravity}}, \tag{18.18}$$

where the minus sign reflects the fact that the potential energy of the ball decreases.

The work done by gravity on the mass is

$$W_{\text{donebygravity}} = -F(y_{\text{f}} - y_{\text{i}}), \tag{18.19}$$

where F is the force due to gravity, and y_{i} and y_{f} are the initial and final positions of the ball, respectively. The negative sign is because gravity points down, which we consider to be the negative direction. For the *constant* gravitational field near Earth's surface, $F = mg$. The change in gravitational potential energy of the mass is

$$-\Delta U_{\text{G}} = W_{\text{donebygravity}} = -F(y_{\text{f}} - y_{\text{i}}) = -mg(y_{\text{f}} - y_{\text{i}}), \text{ or } \Delta U_{\text{G}} = mg(y_{\text{f}} - y_{\text{i}}). \tag{18.20}$$

Note that $y_{\text{f}} - y_{\text{i}}$ is just the negative of the height h from which the mass falls, so we usually just write $\Delta U_{\text{G}} = -mgh$.

We now apply the same reasoning to a charge in an electric field to find the electric potential energy. The change ΔU_{E} in electric potential energy is the work done by the electric field to move a charge q from an initial position x_{i} to a final position x_{f} ($-\Delta U_{\text{E}} = W_{\text{donebyE-field}}$). The definition of work does not change, except that now the work is done by the electric field: $W_{\text{donebyE-field}} = F(x_{\text{f}} - x_{\text{i}})$. For a charge that falls through a *constant* electric field E, the force applied to the charge by the electric field is $F = qE$. The change in electric potential energy of the charge is thus

$$-\Delta U_{\text{E}} = W_{\text{donebyE-field}} = Fd = qE(x_{\text{f}} - x_{\text{i}}) \tag{18.21}$$

or

$$\Delta U_{\text{E}} = -qE(x_{\text{f}} - x_{\text{i}}). \tag{18.22}$$

This equation gives the change in electric potential energy of a charge q when it moves from position x_{i} to position x_{f} in a *constant* electric field E.

Figure 18.22 shows how this analogy would work if we were close to Earth's surface, where gravity is constant. The top image shows a charge accelerating due to a constant electric field. Likewise, the round mass in the bottom image accelerates due to a constant gravitation field. In both cases, the potential energy of the particle decreases, and its kinetic energy increases.

Figure 18.22 In the top picture, a mass accelerates due to a constant electric field. In the bottom picture, the mass accelerates due to a constant gravitational field.

WATCH PHYSICS

Analogy between Gravity and Electricity

This video discusses the analogy between gravitational potential energy and electric potential energy. It reviews the concepts of work and potential energy and shows the connection between a mass in a uniform gravitation field, such as on Earth's surface, and an electric charge in a uniform electric field.

Click to view content (https://www.openstax.org/l/28grav-elec)

If the electric field is not constant, then the equation $\Delta U_{\mathrm{E}} = -qE\,(x_{\mathrm{f}} - x_{\mathrm{i}})$ is not valid, and deriving the electric potential energy becomes more involved. For example, consider the electric potential energy of an assembly of two point charges q_1 and q_2 of the same sign that are initially very far apart. We start by placing charge q_1 at the origin of our coordinate system. This takes no electrical energy, because there is no electric field at the origin (because charge q_2 is very far away). We then bring charge q_2 in from very far away to a distance r from the center of charge q_1 . This requires some effort, because the electric field of charge q_1 applies a repulsive force on charge q_2 . The energy it takes to assemble these two charges can be recuperated if we let them fly apart again. Thus, the charges have potential energy when they are a distance r apart. It turns out that the electric potential energy of a pair of point charges q_1 and q_2 a distance r apart is

$$U_{\mathrm{E}} = \frac{kq_1q_2}{r} \tag{18.23}$$

To recap, if charges q_1 and q_2 are free to move, they can accumulate kinetic energy by flying apart, and this kinetic energy can be used to do work. The maximum amount of work the two charges can do (if they fly infinitely far from each other) is given by the equation above.

Notice that if the two charges have opposite signs, then the potential energy is negative. This means that the charges have more potential to do work when they are *far* apart than when they are at a distance r apart. This makes sense: Opposite charges attract, so the charges can gain more kinetic energy if they attract each other from far away than if they start at only a short distance apart. Thus, they have more potential to do work when they are far apart. Figure 18.23 summarizes how the electric potential energy depends on charge and separation.

Figure 18.23 The potential energy depends on the sign of the charges and their separation. The arrows on the charges indicate the direction in which the charges would move if released. When charges with the same sign are far apart, their potential energy is low, as shown in the top panel for two positive charges. The situation is the reverse for charges of opposite signs, as shown in the bottom panel.

Electric Potential

Recall that to find the force applied by a fixed charge Q on any arbitrary test charge q, it was convenient to define the electric field, which is the force per unit charge applied by Q on any test charge that we place in its electric field. The same strategy is used here with electric potential energy: We now define the **electric potential** V, which is the electric potential energy per unit charge.

$$V = \frac{U_E}{q} \tag{18.24}$$

Normally, the electric potential is simply called the *potential* or *voltage*. The units for the potential are J/C, which are given the name *volt* (V) after the Italian physicist Alessandro Volta (1745–1827). From the equation $U_E = kq_1q_2/r$, the electric potential a distance r from a point charge q_1 is

$$V = \frac{U_E}{q_2} = \frac{kq_1}{r}. \tag{18.25}$$

This equation gives the energy required per unit charge to bring a charge q_2 from infinity to a distance r from a point charge q_1. Mathematically, this is written as

$$V = \left.\frac{U_E}{q_2}\right|_{R=r} - \left.\frac{U_E}{q_2}\right|_{R=\infty}. \tag{18.26}$$

Note that this equation actually represents a *difference* in electric potential. However, because the second term is zero, it is normally not written, and we speak of the electric potential instead of the electric potential difference, or we just say the potential difference, or voltage). Below, when we consider the electric potential energy per unit charge between two points not infinitely far apart, we speak of electric potential *difference* explicitly. Just remember that electric potential and electric potential difference are really the same thing; the former is used just when the electric potential energy is zero in either the initial or final charge configuration.

Coming back now to the electric potential a distance r from a point charge q_1, note that q_1 can be any arbitrary point charge, so we can drop the subscripts and simply write

$$V = \frac{kq}{r}. \tag{18.27}$$

Now consider the electric potential near a group of charges q_1, q_2, and q_3, as drawn in Figure 18.24. The electric potential is

derived by considering the electric field. Electric fields follow the principle of superposition and can be simply added together, so the electric potential from different charges also add together. Thus, the electric potential of a point near a group of charges is

$$V = \frac{kq_1}{r_1} + \frac{kq_2}{r_2} + \frac{kq_3}{r_3} + \cdots. \tag{18.28}$$

where $r_1, r_2, r_3, \ldots$, are the distances from the center of charges $q_1, q_2, q_3, \ldots$ to the point of interest, as shown in Figure 18.24.

Figure 18.24 The potential at the red point is simply the sum of the potentials due to each individual charge.

Now let's consider the electric potential in a uniform electric field. From the equation $\Delta U_E = -qE(x_f - x_i)$, we see that the potential difference in going from x_i to x_f in a uniform electric field E is

$$\Delta V = \frac{\Delta U_E}{q} = -E(x_f - x_i). \tag{18.29}$$

TIPS FOR SUCCESS

Notice from the equation $\Delta V = -E(x_f - x_i)$ that the electric field can be written as

$$E = \frac{\Delta V}{x_f - x_i} \tag{18.30}$$

which means that the electric field has units of V/m. Thus, if you know the potential difference between two points, calculating the electric field is very simple—you simply divide the potential difference by the distance!

Notice that a positive charge in a region with high potential will experience a force pushing it toward regions of lower potential. In this sense, potential is like pressure for fluids. Imagine a pipe containing fluid, with the fluid at one end of the pipe under high pressure and the fluid at the other end of the pipe under low pressure. If nothing prevents the fluid from flowing, it will flow from the high-pressure end to the low-pressure end. Likewise, a positive charge that is free to move will move from a region with high potential to a region with lower potential.

WATCH PHYSICS

Voltage

This video starts from electric potential energy and explains how this is related to electric potential (or voltage). The lecturer calculates the electric potential created by a uniform electric field.

Click to view content (https://www.openstax.org/l/28voltage)

GRASP CHECK

What is the voltage difference between the positions $x_f = 11\text{ m}$ and $x_i = 5.0\text{ m}$ in an electric field of $(2.0\text{ N/C})\,\hat{x}$?

a. 6 V

b. 12 V
c. 24 V
d. 32 V

LINKS TO PHYSICS

Electric Animals

Many animals generate and/or detect electric fields. This is useful for activities such as hunting, defense, navigation, communication, and mating. Because salt water is a relatively good conductor, electric fish have evolved in all the world's oceans. These fish have intrigued humans since the earliest times. In the nineteenth century, parties were even organized where the main attraction was getting a jolt from an electric fish! Scientists also studied electric fish to learn about electricity. Alessandro Volta based his research that led to batteries in 1799 on electric fish. He even referred to batteries as artificial electric organs, because he saw them as imitations of the electric organs of electric fish.

Animals that generate electricity are called *electrogenic* and those that detect electric fields are called *electroreceptive*. Most fish that are electrogenic are also electroreceptive. One of the most well-known electric fish is the electric eel (see Figure 18.25), which is both electrogenic and electroreceptive. These fish have three pairs of organs that produce the electric charge: the main organ, Hunter's organ, and Sach's organ. Together, these organs account for more than 80percent of the fish's body.

Electric eels can produce electric discharges of much greater voltage than what you would get from a standard wall socket. These discharges can stun or even kill their prey. They also use low-intensity discharges to navigate. The electric fields they generate reflect off nearby obstacles or animals and are then detected by electroreceptors in the eel's skin. The three organs that produce electricity contain electrolytes, which are substances that ionize when dissolved in water (or other liquids). An ionized atom or molecule is one that has lost or gained at least one electron, so it carries a net charge. Thus, a liquid solution containing an electrolyte conducts electricity, because the ions in the solution can move if an electric field is applied.

To produce large discharges, the main organ is used. It contains approximately 6,000 rows of electroplaques connected in a long chain. Connected this way, the voltage between electroplaques adds up, creating a large final voltage. Each electroplaque consists of a column of cells controlled by an excitor nerve. When triggered by the excitor nerve, the electroplaques allow ionized sodium to flow through them, creating a potential difference between electroplaques. These potentials add up, and a large current can flow through the electrolyte.

This geometry is reflected in batteries, which also use stacks of plates to produce larger potential differences.

Figure 18.25 An electric eel in its natural environment. (credit: Steven G. Johnson)

GRASP CHECK

If an electric eel produces 1,000 V, which voltage is produced by each electroplaque in the main organ?

a. 0.17 mV
b. 1.7 mV
c. 17 mV
d. 170 mV

WORKED EXAMPLE

X-ray Tube

Dentists use X-rays to image their patients' teeth and bones. The X-ray tubes that generate X-rays contain an electron source separated by about 10 cm from a metallic target. The electrons are accelerated from the source to the target by a uniform electric field with a magnitude of about 100 kN/C, as drawn in Figure 18.26. When the electrons hit the target, X-rays are produced. (a) What is the potential difference between the electron source and the metallic target? (b) What is the kinetic energy of the electrons when they reach the target, assuming that the electrons start at rest?

Figure 18.26 In an X-ray tube, a large current flows through the electron source, causing electrons to be ejected from the electron source. The ejected electrons are accelerated toward the target by the electric field. When they strike the target, X-rays are produced.

STRATEGY FOR (A)

Use the equation $\Delta V = -E(x_f - x_i)$ to find the potential difference given a constant electric field. Define the source position as $x_i = 0$ and the target position as $x_f = 10\text{ cm}$. To accelerate the electrons in the positive x direction, the electric field must point in the negative x direction. This way, the force $F = qE$ on the electrons will point in the positive x direction, because both q and E are negative. Thus, $E = -100 \times 10^3\text{ N/C}$.

Solution for (a)

Using $x_i = 0$ and $x_f = 10\text{ cm} = 0.10\text{ m}$, the equation $\Delta V = -E(x_f - x_i)$ tells us that the potential difference between the electron source and the target is

$$\Delta V = -E(x_f - x_i) = -\left(-100 \times 10^3\text{ N/C}\right)(0.10\text{ m} - 0) = +10\text{ kV}. \tag{18.31}$$

Discussion for (a)

The potential difference is positive, so the energy per unit positive charge is higher at the target than at the source. This means that free positive charges would fall from the target to the source. However, electrons are negative charges, so they accelerate from the source toward the target, gaining kinetic energy as they go.

STRATEGY FOR (B)

Apply conservation of energy to find the final kinetic energy of the electrons. In going from the source to the target, the change in electric potential energy plus the change in kinetic energy of the electrons must be zero, so $\Delta U_E + \Delta K = 0$. The change in electric potential energy for moving through a constant electric field is given by the equation

$$\Delta U_E = -qE(x_f - x_i),$$

where the electric field is $E = -100 \times 10^3\text{ N/C}$. Because the electrons start at rest, their initial kinetic energy is zero. Thus, the change in kinetic energy is simply their final kinetic energy, so $\Delta K = K_f$.

Solution for (b)

Again $x_i = 0$ and $x_f = 10\text{ cm} = 0.10\text{ m}$. The charge of an electron is $q = -1.602 \times 10^{-19}\text{ C}$. Conservation of energy gives

$$\begin{aligned} \Delta U_E + \Delta K &= 0. \\ -qE(x_f - x_i) + K_f &= 0. \\ K_f &= qE(x_f - x_i). \end{aligned} \tag{18.32}$$

Inserting the known values into the right-hand side of this equation gives

$$\begin{aligned} K_f &= \left(-1.60 \times 10^{-19}\text{C}\right)\left(-100 \times 10^{3}\text{N/C}\right)(0.10\text{ m} - 0) \\ &= 1.6 \times 10^{-15}\text{J}. \end{aligned} \tag{18.33}$$

Discussion for (b)

This is a very small energy. However, electrons are very small, so they are easy to accelerate, and this energy is enough to make an electron go extremely fast. You can find their speed by using the definition of kinetic energy, $K = \frac{1}{2}mv^2$. The result is that the electrons are moving at more than 100 million miles per hour!

WORKED EXAMPLE

Electric Potential Energy of Doorknob and Dust Speck

Consider again the doorknob from the example in the previous section. The doorknob is treated as a spherical conductor with a uniform static charge $q_1 = -1.5\text{ nC}$ on its surface. What is the electric potential energy between the doorknob and a speck of dust carrying a charge $q_2 = 0.20\text{ nC}$ at 1.0 cm from the front surface of the doorknob? The diameter of the doorknob is 5.0 cm.

STRATEGY

As we did in the previous section, we treat the charge as if it were concentrated at the center of the doorknob. Again, as you will be able to validate in later physics classes, we can make this simplification, because the charge is uniformly distributed over the surface of the spherical object. Make a sketch of the situation and define a coordinate system, as shown in the image below. We use $+x$ to indicate the outward direction perpendicular to the door, with $x = 0$ at the center of the doorknob. If the diameter of the doorknob is 5.0 cm, its radius is 2.5 cm. Thus, the speck of dust 1.0 cm from the surface of the doorknob is a distance $r = 2.5\text{ cm} + 1.0\text{ cm} = 3.5\text{ cm}$ from the center of the doorknob. To solve this problem, use the equation $U_E = kq_1q_2/r$.

Solution

The charge on the doorknob is $q_1 = -1.5\text{ nC} = -1.5 \times 10^{-9}\text{C}$, and the charge on the speck of dust is $q_2 = 0.20\text{ nC} = 2.0 \times 10^{-10}\text{C}$. The distance $r = 3.5\text{ cm} = 0.035\text{ m}$. Inserting these values into the equation $U_E = kq_1q_2/r$ gives

$$\begin{aligned} U_E &= \frac{kq_1q_2}{r} \\ &= \frac{\left(8.99 \times 10^{9}\text{N}\cdot\text{m}^2/\text{C}^2\right)\left(-1.5 \times 10^{-9}\text{C}\right)\left(2.0 \times 10^{-10}\text{C}\right)}{(0.035\text{ m})} \\ &= -7.7 \times 10^{-8}\text{J}. \end{aligned} \tag{18.34}$$

Discussion

The energy is negative, which means that the energy will decrease that is, get even *more* negative as the speck of dust approaches the doorknob. This helps explain why dust accumulates on objects that carry a static charge. However, note that insulators normally collect more static charge than conductors, because any charge that accumulates on insulators cannot move about on the insulator to find a way to escape. They must simply wait to be removed by some passing moist speck of dust or other host.

Practice Problems

19. What is the electric potential 10 cm from a –10 nC charge?

a. 9.0×10^2 V
b. 9.0×10^3 V
c. 9.0×10^4 V
d. 9.0×10^5 V

20. An electron accelerates from 0 to 10×10^4 m/s in an electric field. Through what potential difference did the electron travel? The mass of an electron is 9.11×10^{-31} kg, and its charge is -1.60×10^{-19} C.

a. 29 mV
b. 290 mV
c. 2,900 mV
d. 29 V

Check Your Understanding

21. Gravitational potential energy is the $(-10\ \text{N/C})\hat{x}$ potential for two masses to do work by virtue of their positions with respect to each other. What is the analogous definition of electric potential energy?

a. Electric potential energy is the potential for two charges to do work by virtue of their positions with respect to the origin point.
b. Electric potential energy is the potential for two charges to do work by virtue of their positions with respect to infinity.
c. Electric potential energy is the potential for two charges to do work by virtue of their positions with respect to each other.
d. Electric potential energy is the potential for single charges to do work by virtue of their positions with respect to their final positions.

22. A negative charge is 10 m from a positive charge. Where would you have to move the negative charge to increase the potential energy of the system?

a. The negative charge should be moved closer to the positive charge.
b. The negative charge should be moved farther away from the positive charge.
c. The negative charge should be moved to infinity.
d. The negative charge should be placed just next to the positive charge.

18.5 Capacitors and Dielectrics

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Calculate the energy stored in a charged capacitor and the capacitance of a capacitor
- Explain the properties of capacitors and dielectrics

Section Key Terms

capacitor	dielectric

Capacitors

Consider again the X-ray tube discussed in the previous sample problem. How can a uniform electric field be produced? A single positive charge produces an electric field that points away from it, as in . This field is not uniform, because the space between the lines increases as you move away from the charge. However, if we combine a positive and a negative charge, we obtain the electric field shown in (a). Notice that, between the charges, the electric field lines are more equally spaced.

What happens if we place, say, five positive charges in a line across from five negative charges, as in Figure 18.27? Now the region between the lines of charge contains a fairly uniform electric field.

Figure 18.27 The red dots are positive charges, and the blue dots are negative charges. The electric-field direction is shown by the red arrows. Notice that the electric field between the positive and negative dots is fairly uniform.

We can extend this idea even further and into two dimensions by placing two metallic plates face to face and charging one with positive charge and the other with an equal magnitude of negative charge. This can be done by connecting one plate to the positive terminal of a battery and the other plate to the negative terminal, as shown in Figure 18.28. The electric field between these charged plates will be extremely uniform.

Figure 18.28 Two parallel metal plates are charged with opposite charge, by connecting the plates to the opposite terminals of a battery. The magnitude of the charge on each plate is the same.

Let's think about the work required to charge these plates. Before the plates are connected to the battery, they are neutral—that is, they have zero net charge. Placing the first positive charge on the left plate and the first negative charge on the right plate requires very little work, because the plates are neutral, so no opposing charges are present. Now consider placing a second positive charge on the left plate and a second negative charge on the right plate. Because the first two charges repel the new arrivals, a force must be applied to the two new charges over a distance to put them on the plates. This is the definition of work, which means that, compared with the first pair, more work is required to put the second pair of charges on the plates. To place the third positive and negative charges on the plates requires yet more work, and so on. Where does this work come from? The battery! Its chemical potential energy is converted into the work required to separate the positive and negative charges.

Although the battery does work, this work remains within the battery-plate system. Therefore, conservation of energy tells us that, if the potential energy of the battery decreases to separate charges, the energy of another part of the system must increase by the same amount. In fact, the energy from the battery is stored in the electric field between the plates. This idea is analogous to considering that the potential energy of a raised hammer is stored in Earth's gravitational field. If the gravitational field were to disappear, the hammer would have no potential energy. Likewise, if no electric field existed between the plates, no energy would be stored between them.

If we now disconnect the plates from the battery, they will hold the energy. We could connect the plates to a lightbulb, for example, and the lightbulb would light up until this energy was used up. These plates thus have the capacity to store energy. For this reason, an arrangement such as this is called a **capacitor**. A capacitor is an arrangement of objects that, by virtue of their geometry, can store energy an electric field.

Various real capacitors are shown in Figure 18.29. They are usually made from conducting plates or sheets that are separated by

an insulating material. They can be flat or rolled up or have other geometries.

Figure 18.29 Some typical capacitors. (credit: Windell Oskay)

The capacity of a capacitor is defined by its capacitance C, which is given by

$$C = \frac{Q}{V}, \tag{18.35}$$

where Q is the *magnitude* of the charge on each capacitor plate, and V is the potential difference in going from the negative plate to the positive plate. This means that both Q and V are always positive, so the capacitance is always positive. We can see from the equation for capacitance that the units of capacitance are C/V, which are called farads (F) after the nineteenth-century English physicist Michael Faraday.

The equation $C = Q/V$ makes sense: A parallel-plate capacitor (like the one shown in Figure 18.28) the size of a football field could hold a lot of charge without requiring too much work per unit charge to push the charge into the capacitor. Thus, Q would be large, and V would be small, so the capacitance C would be very large. Squeezing the same charge into a capacitor the size of a fingernail would require much more work, so V would be very large, and the capacitance would be much smaller.

Although the equation $C = Q/V$ makes it seem that capacitance depends on voltage, in fact it does not. For a given capacitor, the ratio of the charge stored in the capacitor to the voltage difference between the plates of the capacitor always remains the same. Capacitance is determined by the geometry of the capacitor and the materials that it is made from. For a parallel-plate capacitor with nothing between its plates, the capacitance is given by

$$C_0 = \varepsilon_0 \frac{A}{d}, \tag{18.36}$$

where A is the area of the plates of the capacitor and d is their separation. We use C_0 instead of C, because the capacitor has nothing between its plates (in the next section, we'll see what happens when this is not the case). The constant ε_0, read *epsilon zero* is called the permittivity of free space, and its value is

$$\varepsilon_0 = 8.85 \times 10^{-12} \text{ F/m} \tag{18.37}$$

Coming back to the energy stored in a capacitor, we can ask exactly how much energy a capacitor stores. If a capacitor is charged by putting a voltage V across it for example, by connecting it to a battery with voltage V—the electrical potential energy stored in the capacitor is

$$U_E = \frac{1}{2}CV^2. \tag{18.38}$$

Notice that the form of this equation is similar to that for kinetic energy, $K = \frac{1}{2}mv^2$.

WATCH PHYSICS

Where does Capacitance Come From?

This video shows how capacitance is defined and why it depends only on the geometric properties of the capacitor, not on voltage or charge stored. In so doing, it provides a good review of the concepts of work and electric potential.

Click to view content (https://www.openstax.org/l/28capacitance)

GRASP CHECK

If you increase the distance between the plates of a capacitor, how does the capacitance change?

a. Doubling the distance between capacitor plates will reduce the capacitance four fold.
b. Doubling the distance between capacitor plates will reduce the capacitance two fold.
c. Doubling the distance between capacitor plates will increase the capacitance two times.
d. Doubling the distance between capacitor plates will increase the capacitance four times.

Virtual Physics

Charge your Capacitor

Click to view content (http://www.openstax.org/l/28charge-cap)

For this simulation, choose the tab labeled *Introduction* at the top left of the screen. You are presented with a parallel-plate capacitor connected to a variable-voltage battery. The battery is initially at zero volts, so no charge is on the capacitor. Slide the battery slider up and down to change the battery voltage, and observe the charges that accumulate on the plates. Display the capacitance, top-plate charge, and stored energy as you vary the battery voltage. You can also display the electric-field lines in the capacitor. Finally, probe the voltage between different points in this circuit with the help of the voltmeter, and probe the electric field in the capacitor with the help of the electric-field detector.

GRASP CHECK

True or false— In a capacitor, the stored energy is always positive, regardless of whether the top plate is charged with negative or positive charge.

a. false
b. true

WORKED EXAMPLE

Capacitance and Charge Stored in a Parallel Plate Capacitor

(a) What is the capacitance of a parallel-plate capacitor with metal plates, each of area 1.00 m^2, separated by 0.0010 m? (b) What charge is stored in this capacitor if a voltage of 3.00×10^3 V is applied to it?

STRATEGY FOR (A)

Use the equation $C_0 = \varepsilon_0 \frac{A}{d}$.

Solution for (a)

Entering the given values into this equation for the capacitance of a parallel-plate capacitor yields

$$\begin{aligned} C &= \varepsilon_0 \frac{A}{d} \\ &= \left(8.85 \times 10^{-12}\ \text{F/m}\right) \frac{1.00\text{m}^2}{0.0010\ \text{m}} \\ &= 8.9 \times 10^{-9}\ \text{F} \\ &= 8.9\text{nF}. \end{aligned} \tag{18.39}$$

Discussion for (a)

This small value for the capacitance indicates how difficult it is to make a device with a large capacitance. Special techniques help, such as using very-large-area thin foils placed close together or using a dielectric (to be discussed below).

STRATEGY FOR (B)

Knowing C, find the charge stored by solving the equation $C = Q/V$, for the charge Q.

Solution for (b)

The charge Q on the capacitor is

$$\begin{aligned} Q &= CV \\ &= \left(8.9 \times 10^{-9}\text{F}\right)\left(3.00 \times 10^{3}\text{V}\right) \\ &= 2.7 \times 10^{-5}\text{C}. \end{aligned} \tag{18.40}$$

Discussion for (b)

This charge is only slightly greater than typical static electricity charges. More charge could be stored by using a dielectric between the capacitor plates.

WORKED EXAMPLE

What battery is needed to charge a capacitor?

Your friend provides you with a $10\mu\text{F}$ capacitor. To store $120\mu\text{C}$ on this capacitor, what voltage battery should you buy?

STRATEGY

Use the equation $C = Q/V$ to find the voltage needed to charge the capacitor.

Solution

Solving $C = Q/V$ for the voltage gives $V = Q/C$. Inserting $C = 10\mu\text{F} = 10 \times 10^{-6}\text{F}$ and $Q = 120\mu\text{C} = 120 \times 10^{-6}\text{C}$ gives

$$V = \frac{Q}{C} = \frac{120 \times 10^{-6}\text{C}}{10 \times 10^{-6}\text{F}} = 12\text{ V} \tag{18.41}$$

Discussion

Such a battery should be easy to procure. There is still a question of whether the battery contains enough energy to provide the desired charge. The equation $U_E = \frac{1}{2}CV^2$ allows us to calculate the required energy.

$$U_E = \frac{1}{2}CV^2 = \frac{1}{2}\left(10 \times 10^{-6}\text{F}\right)(12\text{ V})^2 = 72\text{ mJ} \tag{18.42}$$

A typical commercial battery can easily provide this much energy.

Practice Problems

23. What is the voltage on a 35 µF with 25 nC of charge?

a. 8.75×10^{-13} V
b. 0.71×10^{-3} V
c. 1.4×10^{-3} V
d. 1.4×10^{3} V

24. Which voltage is across a 100 µF capacitor that stores 10 J of energy?

a. -4.5×10^{2} V
b. 4.5×10^{2} V
c. $\pm 4.5 \times 10^{2}$ V
d. $\pm 9 \times 10^{2}$ V

Dielectrics

Before working through some sample problems, let's look at what happens if we put an insulating material between the plates of a capacitor that has been charged and then disconnected from the charging battery, as illustrated in Figure 18.30. Because the material is insulating, the charge cannot move through it from one plate to the other, so the charge Q on the capacitor does not change. An electric field exists between the plates of a charged capacitor, so the insulating material becomes polarized, as shown in the lower part of the figure. An electrically insulating material that becomes polarized in an electric field is called a **dielectric**.

Figure 18.30 shows that the negative charge in the molecules in the material shifts to the left, toward the positive charge of the capacitor. This shift is due to the electric field, which applies a force to the left on the electrons in the molecules of the dielectric. The right sides of the molecules are now missing a bit of negative charge, so their net charge is positive.

Figure 18.30 The top and bottom capacitors carry the same charge *Q*. The top capacitor has no dielectric between its plates. The bottom capacitor has a dielectric between its plates. The molecules in the dielectric are polarized by the electric field of the capacitor.

All electrically insulating materials are dielectrics, but some are *better* dielectrics than others. A good dielectric is one whose molecules allow their electrons to shift strongly in an electric field. In other words, an electric field pulls their electrons a fair bit away from their atom, but they do not escape completely from their atom (which is why they are insulators).

Figure 18.31 shows a macroscopic view of a dielectric in a charged capacitor. Notice that the electric-field lines in the capacitor with the dielectric are spaced farther apart than the electric-field lines in the capacitor with no dielectric. This means that the electric field in the dielectric is weaker, so it stores less electrical potential energy than the electric field in the capacitor with no dielectric.

Where has this energy gone? In fact, the molecules in the dielectric act like tiny springs, and the energy in the electric field goes into stretching these springs. With the electric field thus weakened, the voltage difference between the two sides of the capacitor is smaller, so it becomes easier to put more charge on the capacitor. Placing a dielectric in a capacitor before charging it therefore allows more charge and potential energy to be stored in the capacitor. A parallel plate with a dielectric has a capacitance of

$$C = \kappa\varepsilon_0 \frac{A}{d} = \kappa C_0, \tag{18.43}$$

where κ (*kappa*) is a dimensionless constant called the *dielectric constant*. Because κ is greater than 1 for dielectrics, the capacitance increases when a dielectric is placed between the capacitor plates. The dielectric constant of several materials is shown in Table 18.1.

Material	Dielectric Constant (κ)
Vacuum	1.00000
Air	1.00059
Fused quartz	3.78
Neoprene rubber	6.7
Nylon	3.4
Paper	3.7
Polystyrene	2.56
Pyrex glass	5.6
Silicon oil	2.5
Strontium titanate	233
Teflon	2.1
Water	80

Table 18.1 Dielectric Constants for Various Materials at 20 °C

Figure 18.31 The top and bottom capacitors carry the same charge Q. The top capacitor has no dielectric between its plates. The bottom capacitor has a dielectric between its plates. Because some electric-field lines terminate and start on polarization charges in the dielectric, the electric field is less strong in the capacitor. Thus, for the same charge, a capacitor stores less energy when it contains a dielectric.

WORKED EXAMPLE

Capacitor for Camera Flash

A typical flash for a point-and-shoot camera uses a capacitor of about $200\ \mu\text{F}$. (a) If the potential difference between the capacitor plates is 100 V—that is, 100 V is placed "across the capacitor," how much energy is stored in the capacitor? (b) If the dielectric used in the capacitor were a 0.010-mm-thick sheet of nylon, what would be the surface area of the capacitor plates?

STRATEGY FOR (A)

Given that $V = 100\ \text{V}$ and $C = 200 \times 10^{-6}\text{F}$, we can use the equation $U_E = \frac{1}{2}CV^2$, to find the electric potential energy stored in the capacitor.

Solution for (a)

Inserting the given quantities into $U_E = \frac{1}{2}CV^2$ gives

$$\begin{aligned} U_E &= \frac{1}{2}CV^2 \\ &= \frac{1}{2}\left(200 \times 10^{-6}\text{F}\right)(100\text{ V})^2 \\ &= 1.0\text{ J}. \end{aligned} \tag{18.44}$$

Discussion for (a)

This is enough energy to lift a 1-kg ball about 1 m up from the ground. The flash lasts for about 0.001 s, so the power delivered by the capacitor during this brief time is $P = \frac{U_E}{t} = \frac{1.0\text{ J}}{0.001\text{ s}} = 1\text{ kW}$. Considering that a car engine delivers about 100 kW of power, this is not bad for a little capacitor!

STRATEGY FOR (B)

Because the capacitor plates are in contact with the dielectric, we know that the spacing between the capacitor plates is $d = 0.010\text{ mm} = 1.0 \times 10^{-5}\text{m}$. From the previous table, the dielectric constant of nylon is $\kappa = 3.4$. We can now use the equation $C = \kappa\varepsilon_0 \frac{A}{d}$ to find the area A of the capacitor.

Solution (b)

Solving the equation for the area A and inserting the known quantities gives

$$\begin{aligned} C &= \kappa\varepsilon_0 \frac{A}{d} \\ A &= \frac{Cd}{\kappa\varepsilon_0} \\ &= \frac{\left(200 \times 10^{-6}\text{F}\right)\left(1.0 \times 10^{-5}\text{m}\right)}{(3.4)\left(8.85 \times 10^{-12}\text{ F/m}\right)} \\ &= 66\text{m}^2. \end{aligned} \tag{18.45}$$

Discussion for (b)

This is much too large an area to roll into a capacitor small enough to fit in a handheld camera. This is why these capacitors don't use simple dielectrics but a more advanced technology to obtain a high capacitance.

Practice Problems

25. With 12 V across a capacitor, it accepts 10 mC of charge. What is its capacitance?

a. 0.83 μF
b. 83 μF
c. 120 μF
d. 830 μF

26. A parallel-plate capacitor has an area of 10 cm^2 and the plates are separated by 100 μm. If the capacitor contains paper between the plates, what is its capacitance?

a. 3.3×10^{-10} F
b. 3.3×10^{-8} F
c. 3.3×10^{-6} F
d. 3.3×10^{-4} F

Check Your Understanding

27. If the area of a parallel-plate capacitor doubles, how is the capacitance affected?

a. The capacitance will remain same.
b. The capacitance will double.
c. The capacitance will increase four times.
d. The capacitance will increase eight times.

28. If you double the area of a parallel-plate capacitor and reduce the distance between the plates by a factor of four, how is the capacitance affected?

a. It will increase by a factor of two.
b. It will increase by a factor of four.
c. It will increase by a factor of six.
d. It will increase by a factor of eight.

KEY TERMS

capacitor arrangement of objects that can store electrical energy by virtue of their geometry

conductor material through which electric charge can easily move, such as metals

Coulomb's law describes the electrostatic force between charged objects, which is proportional to the charge on each object and inversely proportional to the square of the distance between the objects

dielectric electrically insulating material that becomes polarized in an electric field

electric field defines the force per unit charge at all locations in space around a charge distribution

electric potential the electric potential energy per unit charge

electric potential energy the work that a charge can do by virtue of its position in an electric field

electron subatomic particle that carries one indivisible unit of negative electric charge

induction creating an unbalanced charge distribution in an object by moving a charged object toward it (but without touching)

insulator material through which a charge does not move, such as rubber

inverse-square law law that has the form of a ratio, with the denominator being the distance squared

law of conservation of charge states that total charge is constant in any process

polarization separation of charge induced by nearby excess charge

proton subatomic particle that carries the same magnitude charge as the electron, but its charge is positive

test charge positive electric charge whose with a charge magnitude so small that it does not significantly perturb any nearby charge distribution

SECTION SUMMARY

18.1 Electrical Charges, Conservation of Charge, and Transfer of Charge

- Electric charge is a conserved quantity, which means it can be neither created nor destroyed.
- Electric charge comes in two varieties, which are called *positive* and *negative*.
- Charges with the same sign repel each other. Charges with opposite signs attract each other.
- Charges can move easily in conducting material. Charges cannot move easily in an insulating material.
- Objects can be charged in three ways: by contact, by conduction, and by induction.
- Although a polarized object may be neutral, its electrical charge is unbalanced, so one side of the object has excess negative charge and the other side has an equal magnitude of excess positive charge.

18.2 Coulomb's law

- Coulomb's law is an inverse square law and describes the electrostatic force between particles.
- The electrostatic force between charged objects is proportional to the charge on each object and inversely proportional to the distance squared between the objects.
- If Coulomb's law gives a negative result, the force is attractive; if the result is positive, the force is repulsive.

18.3 Electric Field

- The electric field defines the force per unit charge in the space around a charge distribution.
- For a point charge or a sphere of uniform charge, the electric field is inversely proportional to the distance from the point charge or from the center of the sphere.
- Electric-field lines never cross each other.
- More force is applied to a charge in a region with many electric field lines than in a region with few electric field lines.
- Electric field lines start at positive charges and point away from positive charges. They end at negative charges and point toward negative charges.

18.4 Electric Potential

- Electric potential energy is a concept similar to gravitational potential energy: It is the potential that charges have to do work by virtue of their positions relative to each other.
- Electric potential is the electric potential energy per unit charge.
- The potential is always measured between two points, where one point may be at infinity.
- Positive charges move from regions of high potential to regions of low potential.
- Negative charges move from regions of low potential to regions of high potential.

18.5 Capacitors and Dielectrics

- The capacitance of a capacitor depends only on the geometry of the capacitor and the materials from which it is made. It does not depend on the voltage across the capacitor.
- Capacitors store electrical energy in the electric field between their plates.

- A dielectric material is an insulator that is polarized in an electric field.
- Putting a dielectric between the plates of a capacitor increases the capacitance of the capacitor.

KEY EQUATIONS

18.2 Coulomb's law

Coulomb's law	$F = \frac{kq_1q_2}{r^2}$

18.3 Electric Field

electric field	$\vec{E} = \frac{\vec{F}}{q_{\text{test}}}$
magnitude of electric field of point charge	$E = \frac{k\,\lvert Q\rvert}{r^2}$

18.4 Electric Potential

change in electric potential energy for a charge that moves in a constant electric field	$\Delta U_{\text{E}} = -qE\,(x_{\text{f}} - x_{\text{i}})$
electric potential energy of a charge a distance r from a point charge or sphere of uniform charge	$U_{\text{E}} = \frac{kQq}{r}$
definition of electric potential	$V = \frac{U_{\text{E}}}{q}$
change in electric potential for a charge that moves in a constant electric field	$\Delta V = -E\,(x_{\text{f}} - x_{\text{i}})$
electric potential of a charge a distance r from a point charge or sphere of uniform charge	$V = \frac{kQ}{r}$

18.5 Capacitors and Dielectrics

capacitance	$C = \frac{Q}{V}$
energy stored in a capacitor	$U_E = \frac{1}{2}CV^2$
capacitance of a parallel-plate capacitor	$C = \kappa\varepsilon_0\frac{A}{d}$

CHAPTER REVIEW

Concept Items

18.1 Electrical Charges, Conservation of Charge, and Transfer of Charge

1. There are very large numbers of charged particles in most objects. Why, then, don't most objects exhibit static electric effects?
   a. Most objects are neutral.
   b. Most objects have positive charge only.
   c. Most objects have negative charge only.
   d. Most objects have excess protons.

2. Can an insulating material be used to charge a conductor? If so, how? If not, why not?
   a. No, an insulator cannot charge a conductor by induction.
   b. No, an insulating material cannot charge a conductor.
   c. Yes, an uncharged insulator can charge a conductor by induction.
   d. Yes, a charged insulator can charge a conductor upon contact.

3. True or false—A liquid can be an insulating material.
   a. true
   b. false

18.2 Coulomb's law

4. Two plastic spheres with uniform charge repel each other with a force of 10 N . If you remove the charge from one sphere, what will be the force between the spheres?
   a. The force will be 15 N.
   b. The force will be 10 N.
   c. The force will be 5 N.
   d. The force will be zero.

5. What creates a greater magnitude of force, two charges

$+q$ a distance r apart or two charges $-q$ the same distance apart?

a. Two charges $+q$ a distance r away
b. Two charges $-q$ a distance r away
c. The magnitudes of forces are equal.

6. In Newton's law of universal gravitation, the force between two masses is proportional to the product of the two masses. What plays the role of mass in Coulomb's law?

a. the electric charge
b. the electric dipole
c. the electric monopole
d. the electric quadruple

18.3 Electric Field

7. Why can electric fields not cross each other?

a. Many electric-field lines can exist at any given point in space.
b. No electric-field lines can exist at any given point in space.
c. Only a single electric-field line can exist at any given point in space.
d. Two electric-field lines can exist at the same point in space.

8. A constant electric field is $(4.5 \times 10^5\ \text{N/C})\hat{y}$. In which direction is the force on a –20 nC charge placed in this field?

a. The direction of the force is in the $+\hat{x}$ direction.
b. The direction of the force is in the $+\hat{x}$ direction.
c. The direction of the force is in the $-\hat{y}$ direction.
d. The direction of the force is in the $+\hat{y}$ direction.

18.4 Electric Potential

9. True or false—The potential from a group of charges is the sum of the potentials from each individual charge.

a. false
b. true

10. True or false—The characteristics of an electric field make it analogous to the gravitational field near the surface of Earth.

a. false
b. true

11. An electron moves in an electric field. Does it move toward regions of higher potential or lower potential? Explain.

a. It moves toward regions of higher potential because its charge is negative.
b. It moves toward regions of lower potential because its charge is negative
c. It moves toward regions of higher potential because its charge is positive.
d. It moves toward regions of lower potential because its charge is positive.

18.5 Capacitors and Dielectrics

12. You insert a dielectric into an air-filled capacitor. How does this affect the energy stored in the capacitor?

a. Energy stored in the capacitor will remain same.
b. Energy stored in the capacitor will decrease.
c. Energy stored in the capacitor will increase.
d. Energy stored in the capacitor will increase first, and then it will decrease.

13. True or false— Placing a dielectric between the plates of a capacitor increases the energy of the capacitor.

a. false
b. true

14. True or false— The electric field in an air-filled capacitor is reduced when a dielectric is inserted between the plates.

a. false
b. true

Critical Thinking Items

18.1 Electrical Charges, Conservation of Charge, and Transfer of Charge

15. If you dive into a pool of seawater through which an equal amount of positively and negatively charged particles is moving, will you receive an electric shock?

a. Yes, because negatively charged particles are moving.
b. No, because positively charged particles are moving.
c. Yes, because positively and negatively charged particles are moving.
d. No, because equal amounts of positively and negatively charged particles are moving.

16. True or false—The high-voltage wires that you see connected to tall metal-frame towers are held aloft by insulating connectors, and these wires are wrapped in an insulating material.

a. true
b. false

17. By considering the molecules of an insulator, explain how an insulator can be overall neutral but carry a surface charge when polarized.

a. Inside the insulator, the oppositely charged ends of

the molecules cancel each other.
b. Inside the insulator, the oppositely charged ends of the molecules do not cancel each other.
c. The electron distribution in all the molecules shifts in every possible direction, leaving an excess of positive charge on the opposite end of each molecule.
d. The electron distribution in all the molecules shifts in a given direction, leaving an excess of negative charge on the opposite end of each molecule.

18.2 Coulomb's law

18. In terms of Coulomb's law, why are water molecules attracted by positive and negative charges?
a. Water molecules are neutral.
b. Water molecules have a third type of charge that is attracted by positive as well as negative charges.
c. Water molecules are polar.
d. Water molecule have either an excess of electrons or an excess of protons.

19. A negative lightning strike occurs when a negatively charged cloud discharges its excess electrons to the positively charged ground. If you observe a cloud-to-cloud lightning strike, what can you say about the charge on the area of the cloud struck by lightning?
a. The area of the cloud that was struck by lightning had a positive charge.
b. The area of the cloud that was struck by lightning had a negative charge.
c. The area of the cloud that was struck by lightning is neutral.
d. The area of the cloud that was struck by lightning had a third type of charge.

18.3 Electric Field

20. An arbitrary electric field passes through a box-shaped volume. There are no charges in the box. If 11 electric-field lines enter the box, how many electric-field lines must exit the box?
a. nine electric field lines
b. 10 electric field lines
c. 11 electric field lines
d. 12 electric field lines

21. In a science-fiction movie, a villain emits a radial electric field to repulse the hero. Knowing that the hero is electrically neutral, is this possible? Explain your reasoning.
a. No, because an electrically neutral body cannot be repelled or attracted.
b. No, because an electrically neutral body can be attracted but not repelled.
c. Yes, because an electrically neutral body can be repelled or attracted.
d. Yes, because an electrically neutral body can be repelled.

18.4 Electric Potential

22. What is the relationship between voltage and energy? More precisely, what is the relationship between potential difference and electric potential?
a. Voltage is the energy per unit mass at some point in space.
b. Voltage is the energy per unit length in space.
c. Voltage is the energy per unit charge at some point in space.
d. Voltage is the energy per unit area in space.

23. Three parallel plates are stacked above each other, with a separation between each plate. If the potential difference between the first two plates is ΔV_1 and the potential between the second two plates is ΔV_2, what is the potential difference between the first and the third plates?
a. $\Delta V_3 = \Delta V_2 + \Delta V_1$
b. $\Delta V_3 = \Delta V_2 - \Delta V_1$
c. $\Delta V_3 = \Delta V_2 / \Delta V_1$
d. $\Delta V_3 = \Delta V_2 \times \Delta V_1$

18.5 Capacitors and Dielectrics

24. When you insert a dielectric into a capacitor, the energy stored in the capacitor decreases. If you take the dielectric out, the energy increases again. Where does this energy go in the former case, and where does the energy come from in the latter case?
a. Energy is utilized to remove the dielectric and is released when the dielectric is introduced between the plates.
b. Energy is released when the dielectric is added and is utilized when the dielectric is introduced between the plates.
c. Energy is utilized to polarize the dielectric and is released when the dielectric is introduced between the plates.
d. Energy is released to polarize the dielectric and is utilized when dielectric is introduced between the plates.

Problems

18.1 Electrical Charges, Conservation of Charge, and Transfer of Charge

25. A dust particle acquires a charge of –13 nC. How many excess electrons does it carry?
a. 20.8×10^{-28} electrons
b. $20.8 \times {}^{-19}$ electrons
c. 8.1×10^{10} electrons
d. 8.1×10^{19} electrons

26. Two identical conducting spheres are charged with a net charge of +5.0 q on the first sphere and a net charge of –8.0 q on the second sphere. The spheres are brought together, allowed to touch, and then separated. What is the net charge on each sphere now?
a. $-3.0q$
b. $-1.5q$
c. $+1.5q$
d. $+3.0q$

18.2 Coulomb's law

27. Two particles with equal charge experience a force of 10 nN when they are 30 cm apart. What is the magnitude of the charge on each particle?
a. -5.8×10^{-10} C
b. -3.2×10^{-10} C
c. $+3.2 \times 10^{-10}$ C
d. $+1.4 \times 10^{-5}$ C

28. Three charges are on a line. The left charge is $q_1 = 2.0$ nC . The middle charge is $q_2 = 5.0$ nC . The right charge is $q_3 = -3.0$ nC . The left and right charges are 2.0 cm from the middle charge. What is the force on the middle charge?
a. -5.6×10^{-4} N to the left
b. -1.12×10^{-4} N to the left
c. $+1.12 \times 10^{-4}$ N to the right
d. 5.6×10^{-4} N to the right

18.3 Electric Field

29. An electric field $(15\text{ N/C})\hat{z}$ applies a force $(-3 \times 10^{-6}\text{ N})\hat{z}$ on a particle. What is the charge on the particle?
a. -2.0×10^{-7} C
b. 2.0×10^{-7} C
c. 2.0×10^{-8} C
d. 2.0×10^{-9} C

30. Two uniform electric fields are superimposed. The first electric field is $\vec{E}_1 = (14\text{ N/C})\hat{x}$. The second electric field is $\vec{E}_2 = (7.0\text{ N/C})\hat{y}$. With respect to the positive x axis, at which angle will a positive test charge accelerate in this combined field?
a. 27°
b. 54°
c. 90°
d. 108°

18.4 Electric Potential

31. You move a charge q from $r_i = 20$ cm to $r_f = 40$ cm from a fixed charge $Q = 10$ nC. What is the difference in potential for these two positions?
a. -2.2×10^2 V
b. -1.7×10^3 V
c. -2.2×10^4 V
d. -1.7×10^2 V

32. How much work is required from an outside agent to move an electron from $x_i = 0$ to $x_f = 20$ cm in an electric field $(50\text{N/C})\hat{x}$?
a. 1.6×10^{-15} J
b. 1.6×10^{-16} J
c. 1.6×10^{-20} J
d. 1.6×10^{-18} J

18.5 Capacitors and Dielectrics

33. A 4.12 μF parallel-plate capacitor has a plate area of 2,000 cm^2 and a plate separation of 10 μm . What dielectric is between the plates?
a. 1, the dielectric is strontium titanate
b. 466, the dielectric is strontium
c. 699, the dielectric is strontium nitrate
d. 1,000, the dielectric is strontium chloride

34. What is the capacitance of a metal sphere of radius R?
a. $C = \frac{R}{k}$
b. $C = \frac{k}{R}$
c. $C = \frac{V}{Q}$
d. $C = QV$

Performance Task

18.5 Capacitors and Dielectrics

35. Newton's law of universal gravitation is

$$F = \frac{Gm_1 m_2,}{r^2,} \qquad \boxed{18.46}$$

where $G = 6.67 \times 10^{-11}\text{ m}^3/\text{kg} \cdot \text{s}^2$. This describes the gravitational force between two point masses m_1 and m_2.

Coulomb's law is

$$F = \frac{kq_1 q_2}{r^2} \qquad \boxed{18.47}$$

where $k = 8.99 \times 10^9 \text{N} \cdot \text{m}^2/\text{C}^2$. This describes the electric force between two point charges q_1 and q_2.
(a) Describe how the force in each case depends on the distance r between the objects. How do the forces change if the distance is reduced by half? If the distance is doubled?
(b) Describe the similarities and differences between the two laws. Consider the signs of the quantities that create the interaction (i.e., mass and charge), the constants G and k, and their dependence on separation r.
(c) Given that the electric force is much stronger than the gravitational force, discuss why the law for gravitational force was discovered much earlier than the law for electric force.
(d) Consider a hydrogen atom, which is a single proton orbited by a single electron. The electric force holds the electron and proton together so that the hydrogen atom has a radius of about $0.5 \times 10^{-10} \text{m}$. Assuming the force between electron and proton does not change, what would be the approximate radius of the hydrogen atom if $k = 6.67 \times 10^{-11} \text{N} \cdot \text{m}^2/\text{C}^2$?

TEST PREP

Multiple Choice

18.1 Electrical Charges, Conservation of Charge, and Transfer of Charge

36. A neutral hydrogen atom has one proton and one electron. If you remove the electron, what will be the leftover sign of the charge?
a. negative
b. positive
c. zero
d. neutral

37. What is the charge on a proton?
a. $+8.99 \times 10^{-9}$ C
b. -8.99×10^{-9} C
c. $+1.60 \times 10^{-19}$ C
d. -1.60×10^{-19} C

38. True or false—Carbon is more conductive than pure water.
a. true
b. false

39. True or false—Two insulating objects are polarized. To cancel the polarization, it suffices to touch them together.
a. true
b. false

40. How is the charge of the proton related to the charge of the electron?
a. The magnitudes of charge of the proton and the electron are equal, but the charge of the proton is positive, whereas the charge of the electron is negative.
b. The magnitudes of charge of the proton and the electron are unequal, but the charge of the proton is positive, whereas the charge of the electron is negative.
c. The magnitudes of charge of the proton and the electron are equal, but the charge of the proton is negative, whereas the charge of the electron is positive.
d. The magnitudes of charge of the proton and the electron are unequal, but the charge of the proton is negative, whereas the charge of the electron is positive.

18.2 Coulomb's law

41. If you double the distance between two point charges, by which factor does the force between the particles change?
a. 1/2
b. 2
c. 4
d. 1/4

42. The combined charge of all the electrons in a dime is hundreds of thousands of coulombs. Because like charges repel, what keeps the dime from exploding?
a. The dime has an equal number of protons, with positive charge.
b. The dime has more protons than electrons, with positive charge.
c. The dime has fewer protons than electrons, with positive charge.
d. The dime is polarized, with electrons on one side and protons on the other side.

43. How can you modify the charges on two particles to quadruple the force between them without moving them?
a. Increase the distance between the charges by a factor of two.
b. Increase the distance between the charges by a factor of four.
c. Increase the product of the charges by a factor of two
d. Increase the product of the charges by a factor of four.

18.3 Electric Field

44. What is the magnitude of the electric field 12 cm from a charge of 1.5 nC ?
a. 9.4×10^7 N/ C
b. 1.1×10^2 N/C
c. 9.4×10^2 N/C
d. 9.4×10^{-2} N/C

45. A charge distribution has electric field lines pointing into it. What sign is the net charge?
a. positive
b. neutral
c. final
d. negative

46. If five electric field lines come out of point charge q_1 and 10 electric-field lines go into point charge q_2, what is the ratio q_1/q_2?
a. −2
b. −1
c. −1/2
d. 0

47. True or false—The electric-field lines from a positive point charge spread out radially and point outward.
a. false
b. true

18.4 Electric Potential

48. What is the potential at 1.0 m from a point charge $Q = -25$ nC?
a. 6.6×10^2 V
b. -2.3×10^2 V
c. -6.6×10^2 V
d. 2.3×10^2 V

49. Increasing the distance by a factor of two from a point charge will change the potential by a factor of how much?
a. 2
b. 4
c. 1/2
d. 1/4

50. True or false—*Voltage* is the common word for potential difference, because this term is more descriptive than potential difference.
a. false
b. true

18.5 Capacitors and Dielectrics

51. Which magnitude of charge is stored on each plate of a 12 μF capacitor with 12 V applied across it?
a. -1.0×10^{-6} C
b. 1.0×10^{-6} C
c. -1.4×10^{-4} C
d. 1.4×10^{-4} C

52. What is the capacitance of a parallel-plate capacitor with an area of 200 cm^2, a distance of 0.20 mm between the plates, and polystyrene as a dielectric?
a. 2.3 nC
b. 0.89 nC
c. 23 nC
d. 8.9 nC

53. Which factors determine the capacitance of a device?
a. Capacitance depends only on the materials that make up the device.
b. Capacitance depends on the electric field surrounding the device.
c. Capacitance depends on the geometric and material parameters of the device.
d. Capacitance depends only on the mass of the capacitor

Short Answer

18.1 Electrical Charges, Conservation of Charge, and Transfer of Charge

54. Compare the mass of the electron with the mass of the proton.
a. The mass of the electron is about 1,000 times that of the proton.
b. The mass of the proton is about 1,000 times that of the electron.
c. The mass of the electron is about 1,836 times that of the proton.
d. The mass of the proton is about 1,836 times that of the electron.

55. The positive terminal of a battery is connected to one connection of a lightbulb, and the other connection of the lightbulb is connected to the negative terminal of the battery. The battery pushes charge through the circuit but does not become charged itself. Does this violate the law of conservation of charge? Explain.
a. No, because this is a closed circuit.
b. No, because this is an open circuit.
c. Yes, because this is a closed circuit.
d. Yes, because this is an open circuit.

56. Two flat pieces of aluminum foil lay one on top of the other. What happens if you add charge to the top piece of aluminum foil?
a. The charge will distribute over the top of the top

piece.
b. The charge will distribute to the bottom of the bottom piece.
c. The inner surfaces will have excess charge of the opposite sign.
d. The inner surfaces will have excess charge of the same sign.

57. The students in your class count off consecutively so each student has a number. The odd-numbered students are told to act as negative charge, and the even-numbered students are told to act as positive charge. How would you organize them to represent a polarized material?
a. The even-numbered and odd-numbered students will be arranged one after the other.
b. Two even-numbered will be followed by two odd-numbered, and so on.
c. Even-numbered students will be asked to come to the front, whereas odd-numbered students will be asked to go to the back of the class.
d. Half even-numbered and odd-numbered will come to the front, whereas half even-numbered and odd-numbered will go to the back.

58. An ion of iron contains 56 protons. How many electrons must it contain if its net charge is $+5e$?
a. five electrons
b. 51 electrons
c. 56 electrons
d. 61 electrons

59. An insulating rod carries $+2.0\text{ nC}$ of charge. After rubbing it with a material, you find it carries -3 nC of charge. How much charge was transferred to it?
a. -5 nC
b. -3 nC
c. -1 nC
d. $+2.0\text{ nC}$

60. A solid cube carries a charge of $+8e$. You measure the charge on each face of the cube and find that each face carries $+0.5e$ of charge. Is the cube made of conducting or insulating material? Explain.
a. The cube is made of insulating material, because all the charges are on the surface of the cube.
b. The cube is made of conducting material, because some of the charges are inside the cube.
c. The cube is made of insulating material, because all the charges are on the surface of the cube.
d. The cube is made of insulating material, because some of the charges are inside the cube.

61. You have four neutral conducting spheres and a charging device that allows you to place charge q on any neutral object. You want to charge one sphere with a charge $q/2$ and the other three with a charge $q6$. How do you proceed?
a. Charge one sphere with charge q. Touch it simultaneously to the three remaining neutral spheres.
b. Charge one sphere with charge q. Touch it to one other sphere to produce two spheres with charge $\frac{q}{2}$. Touch one of these spheres to one other neutral sphere.
c. Charge one sphere with charge q. Touch it to one other sphere to produce two spheres with charge $\frac{q}{2}$. Touch one of these spheres simultaneously to the two remaining neutral spheres.
d. Charge one sphere with charge q. Touch it simultaneously to two other neutral spheres to produce three spheres with charge $q/3$. Touch one of these spheres to one other neutral sphere.

18.2 Coulomb's law

62. Why does dust stick to the computer screen?
a. The dust is neutral.
b. The dust is polarized.
c. The dust is positively charged.
d. The dust is negatively charged.

63. The force between two charges is 4×10^{-9} N . If the magnitude of one charge is reduced by a factor of two and the distance between the charges is reduced by a factor of two, what is the new force between the charges?
a. 2×10^{-9} N
b. 4×10^{-9} N
c. 6×10^{-9} N
d. 8×10^{-9} N

64. True or false—Coulomb's constant is $k = 8.99 \times 10^9$ $\text{N}\cdot\text{m}^2/\text{C}^2$. Newton's gravitational constant is $G = 6.67 \times 10^{-11}$ $\text{m}^3/\text{kg}\cdot\text{s}^2$. This tells you about the relative strength of the electrostatic force versus that of gravity.
a. true
b. false

65. An atomic nucleus contains 56 protons, for iron. Which force would this nucleus apply on an electron at a distance of 10×10^{-12} m?
a. 0.65×10^{-4} N
b. 0.02×10^{-4} N
c. 1.3×10^{-4} N
d. 72.8×10^{-4} N

18.3 Electric Field

66. The electric field a distance of 10 km from a storm cloud is 1,000 N/C . What is the approximate charge in the

cloud?

a. 0.0011 C
b. 11 C
c. 110 C
d. 1,100 C

67. Which electric field would produce a 10 N force in the $+x$- direction on a charge of – 10 nC ?

a. -1.0×10^9 N/C
b. 1.0×10^9 N/C
c. 1.0×10^{10} N/C
d. 1.0×10^{11} N/C

68. A positive charge is located at $x = 0$. When a negative charge is placed at $x = 10$ cm, what happens to the electric field lines between the charges?

a. The electric field lines become denser between the charges.
b. The electric field lines become denser between the charges.
c. The electric field lines remains same between the charges.
d. The electric field lines will be zero between the charges.

18.4 Electric Potential

69. The energy required to bring a charge $q = -8.8$ nC from far away to 5.5 cm from a point charge Q is 13 mJ. What is the potential at the final position of q?

a. –112 MV
b. –1.5 MV
c. –0.66 MV
d. +1.5 MV

70. How is electric potential related to electric potential energy?

a. Electric potential is the electric potential energy per unit mass at a given position in space.
b. Electric potential is the electric potential energy per unit length at a given position in space. This relation is not dimensionally correct.
c. Electric potential is the electric potential energy per unit area in space.
d. Electric potential is the electric potential energy per unit charge at a given position in space.

71. If it takes 10 mJ to move a charge q from $x_i = 25$ cm to $x_f = -25$ cm in an electric field of $(-20\text{N/C})\hat{x}$, what is the charge q?

a. –1.0 mC
b. +0.25 mC
c. + 1.0 mC
d. +400 mC

72. Given the potential difference between two points and the distance between the points, explain how to obtain the electric field between the points.

a. Add the electric potential to the distance to obtain the electric field.
b. Divide the electric potential by the distance to obtain the electric field.
c. Multiply the electric potential and the distance to obtain the electric field.
d. Subtract the electric potential from the distance to obtain the electric field.

18.5 Capacitors and Dielectrics

73. If you double the voltage across the plates of a capacitor, how is the stored energy affected?

a. Stored energy will decrease two times.
b. Stored energy will decrease four times.
c. Stored energy will increase two times.
d. Stored energy will increase four times.

74. A capacitor with neoprene rubber as the dielectric stores 0.185 mJ of energy with a voltage of 50 V across the plates. If the area of the plates is 500 cm^2, what is the plate separation?

a. 20 μm
b. 20 m
c. 80 μm
d. 80 m

75. Explain why a storm cloud before a lightning strike is like a giant capacitor.

a. The storm cloud acts as a giant charged capacitor, as it can store a large amount of charge.
b. The storm cloud acts as a giant charged capacitor, as it contains a high amount of excess charges.
c. The storm cloud acts as a giant charged capacitor, as it splits in two capacitor plates with equal and opposite charge.
d. The storm cloud acts as a giant charged capacitor, as it splits in two capacitor plates with unequal and opposite charges.

76. A storm cloud is 2 km above the surface of Earth. The lower surface of the cloud is approximately 2 km^2 in area. What is the approximate capacitance of this storm cloud-Earth system?

a. 9×10^{-15} F
b. 9×10^{-9} F
c. 17.7×10^{-15} F
d. 17.7×10^{-9} F

Extended Response

18.1 Electrical Charges, Conservation of Charge, and Transfer of Charge

77. Imagine that the magnitude of the charge on the electron differed very slightly from that of the proton. How would this affect life on Earth and physics in general?

a. Many macroscopic objects would be charged, so we would experience the enormous force of electricity on a daily basis.
b. Many macroscopic objects would be charged, so we would experience the small force of electricity on a daily basis.
c. Many macroscopic objects would be charged, but it would not affect life on Earth and physics in general.
d. Macroscopic objects would remain neutral, so it would not affect life on Earth and physics in general.

78. True or false—Conservation of charge is like balancing a budget.

a. true
b. false

79. True or false—Although wood is an insulator, lightning can travel through a tree to reach Earth.

a. true
b. false

80. True or false—An eccentric inventor attempts to levitate by first placing a large negative charge on himself and then putting a large positive charge on the ceiling of his workshop. Instead, while he attempts to place a large negative charge on himself, his clothes fly off.

a. true
b. false

18.2 Coulomb's law

81. Electrostatic forces are enormous compared to gravitational force. Why do you not notice electrostatic forces in everyday life, whereas you do notice the force due to gravity?

a. Because there are two types of charge, but only one type of mass exists.
b. Because there is only one type of charge, but two types of mass exist.
c. Because opposite charges cancel each other, while gravity does not cancel out.
d. Because opposite charges do not cancel each other, while gravity cancels out.

82. A small metal sphere with a net charge of 3.0 nC is touched to a second small metal sphere that is initially neutral. The spheres are then placed 20 cm apart. What is the force between the spheres?

a. 1.02×10^{-7} N
b. 2.55×10^{-7} N
c. 5.1×10^{-7} N
d. 20.4×10^{-7} N

18.3 Electric Field

83. Point charges are located at each corner of a square with sides of 5.0 cm . The top-left charge is $q_1 = 8.0$ nC The top right charge is $q_2 = 4.0$ nC. The bottom-right charge is $q_3 = 4.0$ nC. The bottom-left charge is $q_4 = 8.0$ nC. What is the electric field at the point midway between charges q_2 and q_3?

a. $(-2.1 \times 10^4 \text{ N/C})\hat{x}$
b. $(2.3 \times 10^4 \text{ N/C})\hat{x}$
c. $(4.1 \times 10^4 \text{ N/C})\hat{x}$
d. $(4.6 \times 10^4 \text{ N/C})\hat{x}$

84. A long straight wire carries a uniform positive charge distribution. Draw the electric field lines in a plane containing the wire at a location far from the ends of the wire. Do not worry about the magnitude of the charge on the wire.

a. Take the wire on the x-axis, and draw electric-field lines perpendicular to it.
b. Take the wire on the *x*-axis, and draw electric-field lines parallel to it.
c. Take the wire on the y-axis, and draw electric-field lines along it.
d. Take the wire on the z-axis, and draw electric-field lines along it.

18.4 Electric Potential

85. A square grid has charges of $Q = 10$ nC are each corner. The sides of the square at 10 cm . How much energy does it require to bring a $q = 1.0$ nC charge from very far away to the point at the center of this square?

a. 1.3×10^{-6} J
b. 2.5×10^{-6} J
c. 3.8×10^{-6} J
d. 5.1×10^{-6} J

86. How are potential difference and electric-field strength related for a constant electric field?

a. The magnitude of electric-field strength is equivalent to the potential divided by the distance.
b. The magnitude of electric-field strength is equivalent to the product of the electric potential and the distance.
c. The magnitude of electric-field strength is

equivalent to the difference between magnitude of the electric potential and the distance.

d. The magnitude of electric-field strength is equivalent to the sum of the magnitude of the electric potential and the distance.

18.5 Capacitors and Dielectrics

87. A 12 μF air-filled capacitor has 12 V across it. If the surface charge on each capacitor plate is $\sigma = 7.2$ mC / m^2, what is the attractive force of one capacitor plate toward the other?

a. 0.81×10^5 N
b. 0.81×10^6 N
c. 1.2×10^5 N
d. 1.2×10^6 N

88. Explain why capacitance should be inversely proportional to the separation between the plates of a capacitor.

a. Capacitance is directly proportional to the electric field, which is inversely proportional to the distance between the capacitor plates.
b. Capacitance is inversely proportional to the electric field, which is inversely proportional to the distance between the capacitor plates.
c. Capacitance is inversely proportional to the electric field, which is directly proportional to the distance between the capacitor plates.
d. Capacitance is directly proportional to the electric field, which is directly proportional to the distance between the capacitor plates.

CHAPTER 19
Electrical Circuits

Figure 19.1 Electric energy in massive quantities is transmitted from this hydroelectric facility, the Srisailam power station located along the Krishna River in India, by the movement of charge—that is, by electric current. (credit: Chintohere, Wikimedia Commons)

Chapter Outline

19.1 Ohm's law

19.2 Series Circuits

19.3 Parallel Circuits

19.4 Electric Power

INTRODUCTION The flicker of numbers on a handheld calculator, nerve impulses carrying signals of vision to the brain, an ultrasound device sending a signal to a computer screen, the brain sending a message for a baby to twitch its toes, an electric train pulling into a station, a hydroelectric plant sending energy to metropolitan and rural users—these and many other examples of electricity involve electric current, which is the movement of charge. Humanity has harnessed electricity, the basis of this technology, to improve our quality of life. Whereas the previous chapter concentrated on static electricity and the fundamental force underlying its behavior, the next two chapters will be devoted to electric and magnetic phenomena involving current. In addition to exploring applications of electricity, we shall gain new insights into the workings of nature.

19.1 Ohm's law

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Describe how current is related to charge and time, and distinguish between direct current and alternating current
- Define resistance and verbally describe Ohm's law
- Calculate current and solve problems involving Ohm's law

Section Key Terms

alternating current	ampere	conventional current	direct current	electric current
nonohmic	ohmic	Ohm's law	resistance	

Direct and Alternating Current

Just as water flows from high to low elevation, electrons that are free to move will travel from a place with low potential to a place with high potential. A battery has two terminals that are at different potentials. If the terminals are connected by a conducting wire, an electric current (charges) will flow, as shown in Figure 19.2. Electrons will then move from the low-potential terminal of the battery (the *negative* end) through the wire and enter the high-potential terminal of the battery (the *positive* end).

Figure 19.2 A battery has a wire connecting the positive and negative terminals, which allows electrons to move from the negative terminal to the positive terminal.

Electric current is the rate at which electric charge moves. A large current, such as that used to start a truck engine, moves a large amount very quickly, whereas a small current, such as that used to operate a hand-held calculator, moves a small amount of charge more slowly. In equation form, electric current I is defined as

$$I = \frac{\Delta Q}{\Delta t}$$

where ΔQ is the amount of charge that flows past a given area and Δt is the time it takes for the charge to move past the area. The SI unit for electric current is the ampere (A), which is named in honor of the French physicist André-Marie Ampère (1775–1836). One **ampere** is one coulomb per second, or

$$1 \text{ A} = 1 \text{ C/s}.$$

Electric current moving through a wire is in many ways similar to water current moving through a pipe. To define the flow of water through a pipe, we can count the water molecules that flow past a given section of the pipe. As shown in Figure 19.3, electric current is very similar. We count the number of electrical charges that flow past a section of a conductor; in this case, a wire.

Figure 19.3 The electric current moving through this wire is the charge that moves past the cross-section A divided by the time it takes for this charge to move past the section *A*.

Assume each particle q in Figure 19.3 carries a charge $q = 1\ \text{nC}$, in which case the total charge shown would be $\Delta Q = 5q = 5\ \text{nC}$. If these charges move past the area A in a time $\Delta t = 1\ \text{ns}$, then the current would be

$$I = \frac{\Delta Q}{\Delta t} = \frac{5\ \text{nC}}{1\ \text{ns}} = 5\ \text{A}. \tag{19.1}$$

Note that we assigned a positive charge to the charges in Figure 19.3. Normally, negative charges—electrons—are the mobile charge in wires, as indicated in Figure 19.2. Positive charges are normally stuck in place in solids and cannot move freely. However, because a positive current moving to the right is the same as a negative current of equal magnitude moving to the left, as shown in Figure 19.4, we define **conventional current** to flow in the direction that a positive charge would flow if it could move. Thus, unless otherwise specified, an electric current is assumed to be composed of positive charges.

Also note that one Coulomb is a significant amount of electric charge, so 5 A is a very large current. Most often you will see current on the order of milliamperes (mA).

Figure 19.4 (a) The electric field points to the right, the current moves to the right, and positive charges move to the right. (b) The equivalent situation but with negative charges moving to the left. The electric field and the current are still to the right.

Snap Lab

Vegetable Current

This lab helps students understand how current works. Given that particles confined in a pipe cannot occupy the same space, pushing more particles into one end of the pipe will force the same number of particles out of the opposite end. This creates a current of particles.

Find a straw and dried peas that can move freely in the straw. Place the straw flat on a table and fill the straw with peas. When you push one pea in at one end, a different pea should come out of the other end. This demonstration is a model for

an electric current. Identify the part of the model that represents electrons and the part of the model that represents the supply of electrical energy. For a period of 30 s, count the number of peas you can push through the straw. When finished, calculate the *pea current* by dividing the number of peas by the time in seconds.

Note that the flow of peas is based on the peas physically bumping into each other; electrons push each other along due to mutually repulsive electrostatic forces.

GRASP CHECK

Suppose four peas per second pass through a straw. If each pea carried a charge of $1\ \text{nC}$, what would the electric current be through the straw?

a. The electric current would be the pea charge multiplied by $1\ \text{nC/pea}$.
b. The electric current would be the pea current calculated in the lab multiplied by $1\ \text{nC/pea}$.
c. The electric current would be the pea current calculated in the lab.
d. The electric current would be the pea charge divided by time.

The direction of conventional current *is the direction that positive charge would flow*. Depending on the situation, positive charges, negative charges, or both may move. In metal wires, as we have seen, current is carried by electrons, so the negative charges move. In ionic solutions, such as salt water, both positively charged and negatively charged ions move. This is also true in nerve cells. Pure positive currents are relatively rare but do occur. History credits American politician and scientist Benjamin Franklin with describing current as the direction that positive charges flow through a wire. He named the type of charge associated with electrons negative long before they were known to carry current in so many situations.

As electrons move through a metal wire, they encounter obstacles such as other electrons, atoms, impurities, etc. The electrons scatter from these obstacles, as depicted in Figure 19.5. Normally, the electrons lose energy with each interaction. [1] To keep the electrons moving thus requires a force, which is supplied by an electric field. The electric field in a wire points from the end of the wire at the higher potential to the end of the wire at the lower potential. Electrons, carrying a negative charge, move on average (or *drift*) in the direction opposite the electric field, as shown in Figure 19.5.

Figure 19.5 Free electrons moving in a conductor make many collisions with other electrons and atoms. The path of one electron is shown. The average velocity of free electrons is in the direction opposite to the electric field. The collisions normally transfer energy to the conductor, so a constant supply of energy is required to maintain a steady current.

So far, we have discussed current that moves constantly in a single direction. This is called **direct current**, because the electric charge flows in only one direction. Direct current is often called *DC* current.

Many sources of electrical power, such as the hydroelectric dam shown at the beginning of this chapter, produce **alternating current**, in which the current direction alternates back and forth. Alternating current is often called *AC current*. Alternating current moves back and forth at regular time intervals, as shown in Figure 19.6. The alternating current that comes from a normal wall socket does not suddenly switch directions. Rather, it increases smoothly up to a maximum current and then smoothly decreases back to zero. It then grows again, but in the opposite direction until it has reached the same maximum value. After that, it decreases smoothly back to zero, and the cycle starts over again.

1This energy is transferred to the wire and becomes thermal energy, which is what makes wires hot when they carry a lot of current.

Figure 19.6 With alternating current, the direction of the current reverses at regular time intervals. The graph on the top shows the current versus time. The negative maxima correspond to the current moving to the left. The positive maxima correspond to current moving to the right. The current alternates regularly and smoothly between these two maxima.

Devices that use AC include vacuum cleaners, fans, power tools, hair dryers, and countless others. These devices obtain the power they require when you plug them into a wall socket. The wall socket is connected to the power grid that provides an alternating potential (AC potential). When your device is plugged in, the AC potential pushes charges back and forth in the circuit of the device, creating an alternating current.

Many devices, however, use DC, such as computers, cell phones, flashlights, and cars. One source of DC is a battery, which provides a constant potential (DC potential) between its terminals. With your device connected to a battery, the DC potential pushes charge in one direction through the circuit of your device, creating a DC current. Another way to produce DC current is by using a transformer, which converts AC potential to DC potential. Small transformers that you can plug into a wall socket are used to charge up your laptop, cell phone, or other electronic device. People generally call this a *charger* or a *battery*, but it is a transformer that transforms AC voltage into DC voltage. The next time someone asks to borrow your laptop charger, tell them that you don't have a laptop charger, but that they may borrow your converter.

WORKED EXAMPLE

Current in a Lightning Strike

A lightning strike can transfer as many as 10^{20} electrons from the cloud to the ground. If the strike lasts 2 ms, what is the average electric current in the lightning?

STRATEGY

Use the definition of current, $I = \frac{\Delta Q}{\Delta t}$. The charge ΔQ from 10^{20} electrons is $\Delta Q = ne$, where $n = 10^{20}$ is the number of electrons and $e = -1.60 \times 10^{-19}\text{C}$ is the charge on the electron. This gives

$$\Delta Q = 10^{20} \times \left(-1.60 \times 10^{-19}\text{C}\right) = -16.0\text{ C}. \tag{19.2}$$

The time $\Delta t = 2 \times 10^{-3}\text{s}$ is the duration of the lightning strike.

Solution

The current in the lightning strike is

$$\begin{aligned} I &= \frac{\Delta Q}{\Delta t} \\ &= \frac{-16.0\text{ C}}{2 \times 10^{-3}\text{s}} \\ &= -8\text{ kA}. \end{aligned} \tag{19.3}$$

Discussion

The negative sign reflects the fact that electrons carry the negative charge. Thus, although the electrons flow from the cloud to the ground, the positive current is defined to flow from the ground to the cloud.

WORKED EXAMPLE

Average Current to Charge a Capacitor

In a circuit containing a capacitor and a resistor, it takes 1 min to charge a 16 μF capacitor by using a 9-V battery. What is the average current during this time?

STRATEGY

We can determine the charge on the capacitor by using the definition of capacitance: $C = \frac{Q}{V}$. When the capacitor is charged by a 9-V battery, the voltage across the capacitor will be $V = 9\text{ V}$. This gives a charge of

$$\begin{aligned} C &= \frac{Q}{V} \\ Q &= CV. \end{aligned} \tag{19.4}$$

By inserting this expression for charge into the equation for current, $I = \frac{\Delta Q}{\Delta t}$, we can find the average current.

Solution

The average current is

$$\begin{aligned} I &= \frac{\Delta Q}{\Delta t} \\ &= \frac{CV}{\Delta t} \\ &= \frac{(16 \times 10^{-6}\text{F})(9\text{ V})}{60\text{ s}} \\ &= 2.4 \times 10^{-6}\text{A} \\ &= 2.4\ \mu\text{A}. \end{aligned} \tag{19.5}$$

Discussion

This small current is typical of the current encountered in circuits such as this.

Practice Problems

1. 10 nC of charge flows through a circuit in 3.0×10^{-6} s. What is the current during this time?
   a. The current passes through the circuit is 3.3×10^{-3} A.
   b. The current passes through the circuit is 30 A.
   c. The current passes through the circuit is 33 A.
   d. The current passes through the circuit is 0.3 A.

2. How long would it take a 10-mA current to charge a capacitor with 5.0 mC?
   a. 0.50 s
   b. 5 ns
   c. 0.50 ns
   d. 50 μs

Resistance and Ohm's Law

As mentioned previously, electrical current in a wire is in many ways similar to water flowing through a pipe. The water current that can flow through a pipe is affected by obstacles in the pipe, such as clogs and narrow sections in the pipe. These obstacles slow down the flow of current through the pipe. Similarly, electrical current in a wire can be slowed down by many factors, including impurities in the metal of the wire or collisions between the charges in the material. These factors create a resistance to the electrical current. **Resistance** is a description of how much a wire or other electrical component opposes the flow of charge through it. In the 19th century, the German physicist Georg Simon Ohm (1787–1854) found experimentally that current through a conductor is proportional to the voltage drop across a current-carrying conductor.

$$I \propto V$$

The constant of proportionality is the resistance R of the material, which leads to

$$V = IR(1.3).$$

This relationship is called **Ohm's law**. It can be viewed as a cause-and-effect relationship, with voltage being the cause and the current being the effect. Ohm's law is an empirical law like that for friction, which means that it is an experimentally observed phenomenon. The units of resistance are volts per ampere, or V/A. We call a V/A an *ohm*, which is represented by the uppercase Greek letter omega (Ω). Thus,

$$1\ \Omega = 1\ \text{V/A}(1.4).$$

Ohm's law holds for most materials and at common temperatures. At very low temperatures, resistance may drop to zero (superconductivity). At very high temperatures, the thermal motion of atoms in the material inhibits the flow of electrons, increasing the resistance. The many substances for which Ohm's law holds are called **ohmic**. Ohmic materials include good conductors like copper, aluminum, and silver, and some poor conductors under certain circumstances. The resistance of ohmic materials remains essentially the same for a wide range of voltage and current.

WATCH PHYSICS

Introduction to Electricity, Circuits, Current, and Resistance

This video presents Ohm's law and shows a simple electrical circuit. The speaker uses the analogy of pressure to describe how electric potential makes charge move. He refers to electric potential as *electric pressure*. Another way of thinking about electric potential is to imagine that lots of particles of the same sign are crowded in a small, confined space. Because these charges have the same sign (they are all positive or all negative), each charge repels the others around it. This means that lots of charges are constantly being pushed towards the outside of the space. A complete electric circuit is like opening a door in the small space: Whichever particles are pushed towards the door now have a way to escape. The higher the electric potential, the harder each particle pushes against the others.

GRASP CHECK

If, instead of a single resistor R, two resistors each with resistance R are drawn in the circuit diagram shown in the video, what can you say about the current through the circuit?

a. The amount of current through the circuit must decrease by half.
b. The amount of current through the circuit must increase by half.
c. The current must remain the same through the circuit.
d. The amount of current through the circuit would be doubled.

Virtual Physics

Ohm's Law

Click to view content (http://www.openstax.org/l/28ohms_law)

This simulation mimics a simple circuit with batteries providing the voltage source and a resistor connected across the batteries. See how the current is affected by modifying the resistance and/or the voltage. Note that the resistance is modeled as an element containing small *scattering centers*. These represent impurities or other obstacles that impede the passage of the current.

GRASP CHECK

In a circuit, if the resistance is left constant and the voltage is doubled (for example, from 3 V to 6 V), how does the current change? Does this conform to Ohm's law?

a. The current will get doubled. This conforms to Ohm's law as the current is proportional to the voltage.
b. The current will double. This does not conform to Ohm's law as the current is proportional to the voltage.

c. The current will increase by half. This conforms to Ohm's law as the current is proportional to the voltage.
d. The current will decrease by half. This does not conform to Ohm's law as the current is proportional to the voltage.

WORKED EXAMPLE

Resistance of a Headlight

What is the resistance of an automobile headlight through which 2.50 A flows when 12.0 V is applied to it?

STRATEGY

Ohm's law tells us $V_{\text{headlight}} = IR_{\text{headlight}}$. The voltage drop in going through the headlight is just the voltage rise supplied by the battery, $V_{\text{headlight}} = V_{\text{battery}}$. We can use this equation and rearrange Ohm's law to find the resistance $R_{\text{headlight}}$ of the headlight.

Solution

Solving Ohm's law for the resistance of the headlight gives

$$\begin{aligned} V_{\text{headlight}} &= IR_{\text{headlight}} \\ V_{\text{battery}} &= IR_{\text{headlight}} \\ R_{\text{headlight}} &= \frac{V_{\text{battery}}}{I} = \frac{12\text{ V}}{2.5\text{ A}} = 4.8\ \Omega. \end{aligned} \tag{19.6}$$

Discussion

This is a relatively small resistance. As we will see below, resistances in circuits are commonly measured in kW or MW.

WORKED EXAMPLE

Determine Resistance from Current-Voltage Graph

Suppose you apply several different voltages across a circuit and measure the current that runs through the circuit. A plot of your results is shown in Figure 19.7. What is the resistance of the circuit?

Figure 19.7 The line shows the current as a function of voltage. Notice that the current is given in milliamperes. For example, at 3 V, the current is 0.003 A, or 3 mA.

STRATEGY

The plot shows that current is proportional to voltage, which is Ohm's law. In Ohm's law ($V = IR$), the constant of proportionality is the resistance R. Because the graph shows current as a function of voltage, we have to rearrange Ohm's law in that form: $I = \frac{V}{R} = \frac{1}{R} \times V$. This shows that the slope of the line of I versus V is $\frac{1}{R}$. Thus, if we find the slope of the line in Figure 19.7, we can calculate the resistance R.

Solution

The slope of the line is the *rise* divided by the *run*. Looking at the lower-left square of the grid, we see that the line rises by 1 mA (0.001 A) and runs over a voltage of 1 V. Thus, the slope of the line is

$$\text{slope} = \frac{0.001\ \text{A}}{1\ \text{V.}} \tag{19.7}$$

Equating the slope with $\frac{1}{R}$ and solving for R gives

$$\begin{aligned} \frac{1}{R} &= \frac{0.001\ \text{A}}{1} \\ R &= \frac{1\ \text{V}}{0.001\ \text{A}} = 1{,}000\ \Omega \end{aligned} \tag{19.8}$$

or 1 k-ohm.

Discussion

This resistance is greater than what we found in the previous example. Resistances such as this are common in electric circuits, as we will discover in the next section. Note that if the line in Figure 19.7 were not straight, then the material would not be ohmic and we would not be able to use Ohm's law. Materials that do not follow Ohm's law are called **nonohmic**.

Practice Problems

3. If you double the voltage across an ohmic resistor, how does the current through the resistor change?
   a. The current will double.
   b. The current will increase by half.
   c. The current will decrease by half.
   d. The current will decrease by a factor of two.

4. The current through a $10\ \Omega$ resistor is $0.025\ \text{A}$. What is the voltage drop across the resistor?
   a. $2.5\ \text{mV}$
   b. $0.25\ \text{V}$
   c. $2.5\ \text{V}$
   d. $0.25\ \text{mV}$

Check Your Understanding

5. What is electric current?
   a. Electric current is the electric charge that is at rest.
   b. Electric current is the electric charge that is moving.
   c. Electric current is the electric charge that moves only from the positive terminal of a battery to the negative terminal.
   d. Electric current is the electric charge that moves only from a region of lower potential to higher potential.

6. What is an ohmic material?
   a. An ohmic material is a material that obeys Ohm's law.
   b. An ohmic material is a material that does not obey Ohm's law.
   c. An ohmic material is a material that has high resistance.
   d. An ohmic material is a material that has low resistance.

7. What is the difference between direct current and alternating current?
   a. Direct current flows continuously in every direction whereas alternating current flows in one direction.
   b. Direct current flows continuously in one direction whereas alternating current reverses its direction at regular time intervals.
   c. Both direct and alternating current flow in one direction but the magnitude of direct current is fixed whereas the magnitude of alternating current changes at regular intervals of time.
   d. Both direct and alternating current changes its direction of flow but the magnitude of direct current is fixed whereas the magnitude of alternating current changes at regular intervals of time.

19.2 Series Circuits

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Interpret circuit diagrams and diagram basic circuit elements
- Calculate equivalent resistance of resistors in series and apply Ohm's law to resistors in series and apply Ohm's law to resistors in series

Section Key Terms

circuit diagram	electric circuit	equivalent resistance
in series	resistor	steady state

Electric Circuits and Resistors

Now that we understand the concept of electric current, let's see what we can do with it. As you are no doubt aware, the modern lifestyle relies heavily on electrical devices. These devices contain ingenious **electric circuits**, which are complete, closed pathways through which electric current flows. Returning to our water analogy, an electric circuit is to electric charge like a network of pipes is to water: The electric circuit guides electric charge from one point to the next, running the charge through various devices along the way to extract work or information.

Electric circuits are made from many materials and cover a huge range of sizes, as shown in Figure 19.8. Computers and cell phones contain electric circuits whose features can be as small as roughly a billionth of a meter (a nanometer, or $10^{-9}\,\text{m}$). The pathways that guide the current in these devices are made by ultraprecise chemical treatments of silicon or other semiconductors. Large power systems, on the other hand, contain electric circuits whose features are on the scale of meters. These systems carry such large electric currents that their physical dimensions must be relatively large.

Figure 19.8 The photo on the left shows a *chip* that contains complex integrated electric circuitry. Chips such as this are at the heart of devices such as computers and cell phones. The photograph on the right shows some typical electric circuitry required for high-power electric power transmission.

The pathways that form electric circuits are made from a conducting material, normally a metal in macroscopic circuits. For example, copper wires inside your school building form the electrical circuits that power lighting, projectors, screens, speakers, etc. To represent an electric circuit, we draw **circuit diagrams**. We use lines and symbols to represent the elements in the circuit. A simple electric circuit diagram is shown on the left side of Figure 19.9. On the right side is an analogous water circuit, which we discuss below.

Figure 19.9 On the left is a circuit diagram showing a battery (in red), a resistor (black zigzag element), and the current *I*. On the right is the analogous water circuit. The pump is like the battery, the sand filter is like the resistor, the water current is like the electrical current, and the reservoir is like the ground.

There are many different symbols that scientists and engineers use in circuit diagrams, but we will focus on four main symbols: the wire, the battery or voltage source, resistors, and the ground. The thin black lines in the electric circuit diagram represent the pathway that the electric charge must follow. These pathways are assumed to be perfect conductors, so electric charge can move along these pathways without losing any energy. In reality, the wires in circuits are not perfect, but they come close enough for our purposes.

The zigzag element labeled *R* is a **resistor**, which is a circuit element that provides a known resistance. Macroscopic resistors are often color coded to indicate their resistance, as shown in Figure 19.10.

The red element in Figure 19.9 is a battery, with its positive and negative terminals indicated; the longer line represents the positive terminal of the battery, and the shorter line represents the negative terminal. Note that the battery icon is not always colored red; this is done in Figure 19.9 just to make it easy to identify.

Finally, the element labeled *ground* on the lower left of the circuit indicates that the circuit is connected to Earth, which is a large, essentially neutral object containing an infinite amount of charge. Among other things, the ground determines the potential of the negative terminal of the battery. Normally, the potential of the ground is defined to be zero: $V_{\text{ground}} \equiv 0$. This

means that the entire lower wire in Figure 19.10 is at a voltage of zero volts.

Figure 19.10 Some typical resistors. The color bands indicate the value of the resistance of each resistor.

The electric current in Figure 19.9 is indicted by the blue line labeled *I*. The arrow indicates the direction in which positive charge would flow in this circuit. Recall that, in metals, electrons are mobile charge carriers, so negative charges actually flow in the opposite direction around this circuit (i.e., counterclockwise). However, we draw the current to show the direction in which positive charge would move.

On the right side of Figure 19.9 is an analogous water circuit. Water at a higher pressure leaves the top of the pump, which is like charges leaving the positive terminal of the battery. The water travels through the pipe, like the charges traveling through the wire. Next, the water goes through a sand filter, which heats up as the water squeezes through. This step is like the charges going through the resistor. When charges flow through a resistor, they do work to heat up the resistor. After flowing through the sand filter, the water has converted its potential energy into heat, so it is at a lower pressure. Likewise, the charges exiting the resistor have converted their potential energy into heat, so they are at a lower voltage. Recall that voltage is just potential energy per charge. Thus, water pressure is analogous to electric potential energy (i.e., voltage). Coming back to the water circuit again, we see that the water returns to the bottom of the pump, which is like the charge returning to the negative terminal of the battery. The water pump uses a source of energy to pump the water back up to a high pressure again, giving it the pressure required to go through the circuit once more. The water pump is like the battery, which uses chemical energy to increase the voltage of the charge up to the level of the positive terminal.

The potential energy per charge at the positive terminal of the battery is the voltage rating of the battery. This voltage is like water pressure in the upper pipe. Just like a higher pressure forces water to move toward a lower pressure, a higher voltage forces electric charge to flow toward a lower voltage. The pump takes water at low pressure and does work on it, ejecting water at a higher pressure. Likewise, a battery takes charge at a low voltage, does work on it, and ejects charge at a higher voltage.

Note that the current in the water circuit of Figure 19.9 is the same throughout the circuit. In other words, if we measured the number of water molecules passing a cross-section of the pipe per unit time at any point in the circuit, we would get the same answer no matter where in the circuit we measured. The same is true of the electrical circuit in the same figure. The electric current is the same at all points in this circuit, including inside the battery and in the resistor. The electric current neither speeds up in the wires nor slows down in the resistor. This would create points where too much or too little charge would be bunched up. Thus, the current is the same at all points in the circuit shown in Figure 19.9.

Although the current is the same everywhere in both the electric and water circuits, the voltage or water pressure changes as you move through the circuits. In the water circuit, the water pressure at the pump outlet stays the same until the water goes through the sand filter, assuming no energy loss in the pipe. Likewise, the voltage in the electrical circuit is the same at all points in a given wire, because we have assumed that the wires are perfect conductors. Thus, as indicated by the constant red color of the upper wire in Figure 19.11, the voltage throughout this wire is constant at $V = V_{\text{battery}}$. The voltage then drops as you go through the resistor, but once you reach the blue wire, the voltage stays at its new level of $V = 0$ all the way to the negative terminal of the battery (i.e., the blue terminal of the battery).

Figure 19.11 The voltage in the red wire is constant at $V = V_{\text{battery}}$ from the positive terminal of the battery to the top of the resistor. The voltage in the blue wire is constant at $V = V_{\text{ground}} = 0$ from the bottom of the resistor to the negative terminal of the battery.

If we go from the blue wire through the battery to the red wire, the voltage increases from $V = 0$ to $V = V_{\text{battery}}$. Likewise, if we go from the blue wire up through the resistor to the red wire, the voltage also goes from $V = 0$ to $V = V_{\text{battery}}$. Thus, using Ohm's law, we can write

$$V_{\text{resistor}} = V_{\text{battery}} = IR.$$

Note that V_{resistor} is measured from the bottom of the resistor to the top, meaning that the top of the resistor is at a higher voltage than the bottom of the resistor. Thus, current flows from the top of the resistor or higher voltage to the bottom of the resistor or lower voltage.

Virtual Physics

Battery-Resistor Circuit

Click to view content (http://www.openstax.org/l/21batteryresist)

Use this simulation to better understand how resistance, voltage, and current are related. The simulation shows a battery with a resistor connected between the terminals of the battery, as in the previous figure. You can modify the battery voltage and the resistance. The simulation shows how electrons react to these changes. It also shows the atomic cores in the resistor and how they are excited and heat up as more current goes through the resistor.

Draw the circuit diagram for the circuit, being sure to draw an arrow indicating the direction of the current. Now pick three spots along the wire. Without changing the settings, allow the simulation to run for 20 s while you count the number of electrons passing through that spot. Record the number on the circuit diagram. Now do the same thing at each of the other two spots in the circuit. What do you notice about the number of charges passing through each spot in 20 s? Remember that that current is defined as the rate that charges flow through the circuit. What does this mean about the current through the entire circuit?

GRASP CHECK

With the voltage slider, give the battery a positive voltage. Notice that the electrons are spaced farther apart in the left wire than they are in the right wire. How does this reflect the voltage in the two wires?

a. The voltage between static charges is directly proportional to the distance between them.
b. The voltage between static charges is directly proportional to square of the distance between them.
c. The voltage between static charges is inversely proportional to the distance between them.
d. The voltage between static charges is inversely proportional to square of the distance between them.

Other possible circuit elements include capacitors and switches. These are drawn as shown on the left side of Figure 19.12. A switch is a device that opens and closes the circuit, like a light switch. It is analogous to a valve in a water circuit, as shown on the right side of Figure 19.12. With the switch open, no current passes through the circuit. With the switch closed, it becomes part of the wire, so the current passes through it with no loss of voltage.

The capacitor is labeled C on the left of Figure 19.12. A capacitor in an electrical circuit is analogous to a flexible membrane in a

water circuit. When the switch is closed in the circuit of Figure 19.12, the battery forces electrical current to flow toward the capacitor, charging the upper capacitor plate with positive charge. As this happens, the voltage across the capacitor plates increases. This is like the membrane in the water circuit: When the valve is opened, the pump forces water to flow toward the membrane, making it stretch to store the excess water. As this happens, the pressure behind the membrane increases.

Now if we open the switch, the capacitor holds the voltage between its plates because the charges have nowhere to go. Likewise, if we close the valve, the water has nowhere to go and the membrane maintains the water pressure in the pipe between itself and the valve.

If the switch is closed for a long time in the electric circuit or if the valve is open for a long time in the water circuit, the current will eventually stop flowing because the capacitor or the membrane will have become completely charged. Each circuit is now in the **steady state**, which means that its characteristics do not change over time. In this case, the steady state is characterized by zero current, and this does not change as long as the switch or valve remains in the same position. In the steady state, no electrical current passes through the capacitor, and no water current passes through the membrane. The voltage difference between the capacitor plates will be the same as the battery voltage. In the water circuit, the pressure behind the membrane will be the same as the pressure created by the pump.

Although the circuit in Figure 19.12 may seem a bit pointless because all that happens when the switch is closed is that the capacitor charges up, it does show the capacitor's ability to store charge. Thus, the capacitor serves as a reservoir for charge. This property of capacitors is used in circuits in many ways. For example, capacitors are used to power circuits while batteries are being charged. In addition, capacitors can serve as filters. To understand this, let's go back to the water analogy. Suppose you have a water hose and are watering your garden. Your friend thinks he's funny, and *kinks* the hose. While the hose is kinked, you experience no water flow. When he lets go, the water starts flowing again. If he does this really fast, you experience water-no water-water-no water, and that's really no way to water your garden. Now imagine that the hose is filling up a big bucket, and you are watering from the bottom of the bucket. As long as you had water in your bucket to begin with and your friend doesn't kink the water hose for too long, you would be able to water your garden without the interruptions. Your friend kinking the water hose is *filtered* by the big bucket's supply of water, so it does not impact your ability to water the garden. We can think of the interruptions in the current (be it water or electrical current) as *noise*. Capacitors act in an analogous way as the water bucket to help filter out the noise. Capacitors have so many uses that it is very rare to find an electronic circuit that does not include some capacitors.

Figure 19.12 On the left is an electrical circuit containing a battery, a switch, and a capacitor. On the left is the analogous water circuit with a pump, a valve, and a stretchable membrane. The pump is like the battery, the valve is like the switch, and the stretchable membrane is like the capacitor. When the switch is closed, electrical current flows as the capacitor charges and its voltage increases. Likewise in the water circuit, when the valve is open, water current flows as the stretchable membrane stretches and the water pressure behind it increases.

WORK IN PHYSICS

What It Takes to be an Electrical Engineer

Physics is used in a wide variety of fields. One field that requires a very thorough knowledge of physics is electrical engineering. An electrical engineer can work on anything from the large-scale power systems that provide power to big cities to the nanoscale electronic circuits that are found in computers and cell phones (Figure 19.13).

In working with power companies, you can be responsible for maintaining the power grid that supplies electrical power to large areas. Although much of this work is done from an office, it is common to be called in for overtime duty after storms or other natural events. Many electrical engineers enjoy this part of the job, which requires them to race around the countryside repairing high-voltage transformers and other equipment. However, one of the more unpleasant aspects of this work is to remove the carcasses of unfortunate squirrels or other animals that have wandered into the transformers.

Other careers in electrical engineering can involve designing circuits for cell phones, which requires cramming some 10 billion transistors into an electronic chip the size of your thumbnail. These jobs can involve much work with computer simulations and can also involve fields other than electronics. For example, the 1-m-diameter lenses that are used to make these circuits (as of 2015) are so precise that they are shipped from the manufacture to the chip fabrication plant in temperature-controlled trucks to ensure that they are held within a certain temperature range. If they heat up or cool down too much, they deform ever so slightly, rendering them useless for the ultrahigh precision photolithography required to manufacture these chips.

In addition to a solid knowledge of physics, electrical engineers must above all be practical. Consider, for example, how one corporation managed to launch some anti-ballistic missiles at the White Sands Missile Test Range in New Mexico in the 1960s. Before launch, the skin of the missile had to be at the same voltage as the rail from which it was launched. The rail was connected to the ground by a large copper wire connected to a stake driven into the sandy earth. The missile, however, was connected by an umbilical cord to the equipment in the control shed a few meters away, which was grounded via a different grounding circuit. Before launching the missile, the voltage difference between the missile skin and the rail had to be less than 2.5 V. After an especially dry spell of weather, the missile could not be launched because the voltage difference stood at 5 V. A group of electrical engineers, including the father of your author, stood around pondering how to reduce the voltage difference. The situation was resolved when one of the engineers realized that urine contains electrolytes and conducts electricity quite well. With that, the four engineers quickly resolved the problem by urinating on the rail spike. The voltage difference immediately dropped to below 2.5 V and the missile was launched on schedule.

Figure 19.13 The systems that electrical engineers work on range from microprocessor circuits (left)] to missile systems (right).

Virtual Physics

Click to view content (http://www.openstax.org/l/21phetcirconstr)

Amuse yourself by building circuits of all different shapes and sizes. This simulation provides you with various standard circuit elements, such as batteries, AC voltage sources, resistors, capacitors, light bulbs, switches, etc. You can connect these in any configuration you like and then see the result.

Build a circuit that starts with a resistor connected to a capacitor. Connect the free side of the resistor to the positive terminal of a battery and the free side of the capacitor to the negative terminal of the battery. Click the *reset dynamics* button to see how the current flows starting with no charge on the capacitor. Now right click on the resistor to change its

value. When you increase the resistance, does the circuit reach the steady state more rapidly or more slowly?

GRASP CHECK

When the circuit has reached the steady state, how does the voltage across the capacitor compare to the voltage of the battery? What is the voltage across the resistor?

a. The voltage across the capacitor is greater than the voltage of the battery. In the steady state, no current flows through this circuit, so the voltage across the resistor is zero.
b. The voltage across the capacitor is smaller than the voltage of the battery. In the steady state, finite current flows through this circuit, so the voltage across the resistor is finite.
c. The voltage across the capacitor is the same as the voltage of the battery. In the steady state, no current flows through this circuit, so the voltage across the resistor is zero.
d. The voltage across the capacitor is the same as the voltage of the battery. In the steady state, finite current flows through this circuit, so the voltage across the resistor is finite.

Resistors in Series and Equivalent Resistance

Now that we have a basic idea of how electrical circuits work, let's see what happens in circuits with more than one circuit element. In this section, we look at resistors in series. Components connected **in series** are connected one after the other in the same branch of a circuit, such as the resistors connected in series on the left side of Figure 19.14.

Figure 19.14 On the left is an electric circuit with three resistors R_1, R_2, and R_3 connected in series. On the right is an electric circuit with one resistor R_{equiv} that is equivalent to the combination of the three resistors R_1, R_2, and R_3.

We will now try to find a single resistance that is equivalent to the three resistors in series on the left side of Figure 19.14. An **equivalent resistor** is a resistor that has the same resistance as the combined resistance of a set of other resistors. In other words, the same current will flow through the left and right circuits in Figure 19.14 if we use the equivalent resistor in the right circuit.

According to Ohm's law, the voltage drop V across a resistor when a current flows through it is $V = IR$ where I is the current in amperes (A) and R is the resistance in ohms (Ω). Another way to think of this is that V is the voltage necessary to make a current I flow through a resistance R. Applying Ohm's law to each resistor on the left circuit of Figure 19.14, we find that the voltage drop across R_1 is $V_1 = IR_1$, that across R_2 is $V_2 = IR_2$, and that across R_3 is $V_3 = IR_3$. The sum of these voltages equals the voltage output of the battery, that is

$$V_{\text{battery}} = V_1 + V_2 + V_3. \tag{19.9}$$

You may wonder why voltages must add up like this. One way to understand this is to go once around the circuit and add up the successive changes in voltage. If you do this around a loop and get back to the starting point, the total change in voltage should be zero, because you end up at the same place that you started. To better understand this, consider the analogy of going for a stroll through some hilly countryside. If you leave your car and walk around, then come back to your car, the total height you gained in your stroll must be the same as the total height you lost, because you end up at the same place as you started. Thus, the gravitational potential energy you gain must be the same as the gravitational potential energy you lose. The same reasoning holds for voltage in going around an electric circuit. Let's apply this reasoning to the left circuit in Figure 19.14. We start just below the battery and move up through the battery, which contributes a voltage *gain* of V_{battery}. Next, we got through the resistors. The voltage *drops* by V_1 in going through resistor R_1, by V_2 in going through resistor R_2, and by V_3 in going through

resistor R_3 . After going through resistor R_3, we arrive back at the starting point, so we add up these four changes in voltage and set the sum equal to zero. This gives

$$0 = V_{\text{battery}} - V_1 - V_2 - V_3. \tag{19.10}$$

which is the same as the previous equation. Note that the minus signs in front of V_1 ,V_2 , and V_3 are because these are voltage *drops*, whereas V_{battery} is a voltage *rise*.

Ohm's law tells us that $V_1 = IR_1$, $V_2 = IR_2$, and $V_3 = IR_3$. Inserting these values into equation $V_{\text{battery}} = V_1 + V_2 + V_3$ gives

$$\begin{aligned} V_{\text{battery}} &= IR_1 + IR_2 + IR_3 \\ &= I(R_1 + R_2 + R_3). \end{aligned} \tag{19.11}$$

Applying this same logic to the right circuit in Figure 19.14 gives

$$V_{\text{battery}} = IR_{\text{equiv.}} \tag{19.12}$$

Dividing the equation $V_{\text{battery}} = I(R_1 + R_2 + R_3)$ by $V_{\text{battery}} = IR_{\text{equiv}}$, we get

$$\begin{aligned} \frac{V_{\text{battery}}}{V_{\text{battery}}} &= \frac{I(R_1+R_2+R_3)}{IR_{\text{equiv}}} \\ R_{\text{equiv}} &= R_1 + R_2 + R_3. \end{aligned} \tag{19.13}$$

This shows that the equivalent resistance for a series of resistors is simply the sum of the resistances of each resistor. In general, N resistors connected in series can be replaced by an equivalent resistor with a resistance of

$$R_{\text{equiv}} = R_1 + R_2 + \cdots + R_N.$$

WATCH PHYSICS

Resistors in Series

This video discusses the basic concepts behind interpreting circuit diagrams and then shows how to calculate the equivalent resistance for resistors in series.

Click to view content (https://www.openstax.org/l/02resistseries)

GRASP CHECK

True or false—In a circuit diagram, we can assume that the voltage is the same at every point in a given wire.

a. false
b. true

WORKED EXAMPLE

Calculation of Equivalent Resistance

In the left circuit of the previous figure, suppose the voltage rating of the battery is 12 V, and the resistances are $R_1 = 1.0\ \Omega$, $R_2 = 6.0\ \Omega$, and $R_3 = 13\ \Omega$. (a) What is the equivalent resistance? (b) What is the current through the circuit?

STRATEGY FOR (A)

Use the equation for the equivalent resistance of resistors connected in series. Because the circuit has three resistances, we only need to keep three terms, so it takes the form

$$R_{\text{equiv}} = R_1 + R_2 + R_3. \tag{19.14}$$

Solution for (a)

Inserting the given resistances into the equation above gives

$$\begin{aligned} R_{\text{equiv}} &= R_1 + R_2 + R_3 \\ &= 1.0\ \Omega + 6.0\ \Omega + 13\ \Omega \\ &= 20\ \Omega. \end{aligned} \tag{19.15}$$

Discussion for (a)

We can thus replace the three resistors R_1, R_2, and R_3 with a single 20- Ω resistor.

STRATEGY FOR (B)

Apply Ohm's law to the circuit on the right side of the previous figure with the equivalent resistor of 20 Ω .

Solution for (b)

The voltage drop across the equivalent resistor must be the same as the voltage rise in the battery. Thus, Ohm's law gives

$$\begin{aligned} V_{\text{battery}} &= IR_{\text{equiv}} \\ I &= \frac{V_{\text{battery}}}{R_{\text{equiv}}} \\ &= \frac{12\ \text{V}}{20\ \Omega} \\ &= 0.60\ \text{A}. \end{aligned} \tag{19.16}$$

Discussion for (b)

To check that this result is reasonable, we calculate the voltage drop across each resistor and verify that they add up to the voltage rating of the battery. The voltage drop across each resistor is

$$\begin{aligned} V_1 &= IR_1 = (0.60\ \text{A})(1.0\Omega) = 0.60\ \text{V} \\ V_2 &= IR_2 = (0.60\ \text{A})(6.0\Omega) = 3.6\ \text{V} \\ V_3 &= IR_3 = (0.60\ \text{A})(13\Omega) = 7.8\ \text{V}. \end{aligned} \tag{19.17}$$

Adding these voltages together gives

$$V_1 + V_2 + V_3 = 0.60\ \text{V} + 3.6\ \text{V} + 7.8\ \text{V} = 12\ \text{V}. \tag{19.18}$$

which is the voltage rating of the battery.

WORKED EXAMPLE

Determine the Unknown Resistance

The circuit shown in figure below contains three resistors of known value and a third element whose resistance R_3 is unknown. Given that the equivalent resistance for the entire circuit is 150 Ω, what is the resistance R_3 ?

STRATEGY

The four resistances in this circuit are connected in series, so we know that they must add up to give the equivalent resistance. We can use this to find the unknown resistance R_3 .

Solution

For four resistances in series, the equation for the equivalent resistance of resistors in series takes the form

$$R_{\text{equiv}} = R_1 + R_2 + R_3 + R_4. \tag{19.19}$$

Solving for R_3 and inserting the known values gives

$$\begin{aligned} R_3 &= R_{\text{equiv}} - R_1 - R_2 - R_4 \\ &= 150\ \Omega - 10\ \Omega - 25\ \Omega - 15\ \Omega \\ &= 100\ \Omega. \end{aligned} \tag{19.20}$$

Discussion

The equivalent resistance of a circuit can be measured with an ohmmeter. This is sometimes useful for determining the effective resistance of elements whose resistance is not marked on the element.

Check your Understanding

8.

Figure 19.15

What circuit element is represented in the figure below?

a. a battery
b. a resistor
c. a capacitor
d. an inductor

9. How would a diagram of two resistors connected in series appear?

a.
b.
c.
d.

19.3 Parallel Circuits

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Interpret circuit diagrams with parallel resistors
- Calculate equivalent resistance of resistor combinations containing series and parallel resistors

Section Key Terms

in parallel

Resistors in Parallel

In the previous section, we learned that resistors in series are resistors that are connected one after the other. If we instead combine resistors by connecting them next to each other, as shown in Figure 19.16, then the resistors are said to be connected *in parallel*. Resistors are **in parallel** when both ends of each resistor are connected directly together.

Note that the tops of the resistors are all connected to the same wire, so the voltage at the top of the each resistor is the same. Likewise, the bottoms of the resistors are all connected to the same wire, so the voltage at the bottom of each resistor is the same. This means that the voltage drop across each resistor is the same. In this case, the voltage drop is the voltage rating V of the battery, because the top and bottom wires connect to the positive and negative terminals of the battery, respectively.

Although the voltage drop across each resistor is the same, we cannot say the same for the current running through each resistor. Thus, I_1, I_2, and I_3 are not necessarily the same, because the resistors R_1, R_2, and R_3 do not necessarily have the same

resistance.

Note that the three resistors in Figure 19.16 provide three different paths through which the current can flow. This means that the equivalent resistance for these three resistors must be less than the smallest of the three resistors. To understand this, imagine that the smallest resistor is the only path through which the current can flow. Now add on the alternate paths by connecting other resistors in parallel. Because the current has more paths to go through, the overall resistance (i.e., the equivalent resistance) will decrease. Therefore, the equivalent resistance must be less than the smallest resistance of the parallel resistors.

Figure 19.16 The left circuit diagram shows three resistors in parallel. The voltage V of the battery is applied across all three resistors. The currents that flow through each branch are not necessarily equal. The right circuit diagram shows an equivalent resistance that replaces the three parallel resistors.

To find the equivalent resistance R_{equiv} of the three resistors R_1, R_2, and R_3, we apply Ohm's law to each resistor. Because the voltage drop across each resistor is V, we obtain

$$V = I_1 R_1, \quad V = I_2 R_2, \quad V = I_3 R_3 \tag{19.21}$$

or

$$I_1 = \frac{V}{R_1}, \quad I_2 = \frac{V}{R_2}, \quad I_3 = \frac{V}{R_3}. \tag{19.22}$$

We also know from conservation of charge that the three currents I_1, I_2, and I_3 must add up to give the current I that goes through the battery. If this were not true, current would have to be mysteriously created or destroyed somewhere in the circuit, which is physically impossible. Thus, we have

$$I = I_1 + I_2 + I_3. \tag{19.23}$$

Inserting the expressions for I_1, I_2, and I_3 into this equation gives

$$I = \frac{V}{R_1} + \frac{V}{R_2} + \frac{V}{R_3} = V\left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}\right) \tag{19.24}$$

or

$$V = I\left(\frac{1}{1/R_1 + 1/R_2 + 1/R_3}\right). \tag{19.25}$$

This formula is just Ohm's law, with the factor in parentheses being the equivalent resistance.

$$V = I\left(\frac{1}{1/R_1 + 1/R_2 + 1/R_3}\right) = IR_{\text{equiv.}} \tag{19.26}$$

Thus, the equivalent resistance for three resistors in parallel is

$$R_{\text{equiv}} = \frac{1}{1/R_1 + 1/R_2 + 1/R_3}. \tag{19.27}$$

The same logic works for any number of resistors in parallel, so the general form of the equation that gives the equivalent resistance of N resistors connected in parallel is

$$R_{\text{equiv}} = \frac{1}{1/R_1 + 1/R_2 + \cdots + 1/R_N}. \tag{19.28}$$

WORKED EXAMPLE

Find the Current through Parallel Resistors

The three circuits below are equivalent. If the voltage rating of the battery is $V_{\text{battery}} = 3\ \text{V}$, what is the equivalent resistance of the circuit and what current runs through the circuit?

STRATEGY

The three resistors are connected in parallel and the voltage drop across them is V_{battery}. Thus, we can apply the equation for the equivalent resistance of resistors in parallel, which takes the form

$$R_{\text{equiv}} = \frac{1}{1/R_1 + 1/R_2 + 1/R_3}. \tag{19.29}$$

The circuit with the equivalent resistance is shown below. Once we know the equivalent resistance, we can use Ohm's law to find the current in the circuit.

Solution

Inserting the given values for the resistance into the equation for equivalent resistance gives

$$\begin{aligned} R_{\text{equiv}} &= \frac{1}{1/R_1+1/R_2+1/R_3} \\ &= \frac{1}{1/10\ \Omega+1/25\ \Omega+1/15\ \Omega} \\ &= 4.84\ \Omega. \end{aligned} \tag{19.30}$$

The current through the circuit is thus

$$\begin{aligned} V &= IR \\ I &= \frac{V}{R} \\ &= \frac{3\ \text{V}}{4.84\ \Omega} \\ &= 0.62\ \text{A}. \end{aligned} \tag{19.31}$$

Discussion

Although 0.62 A flows through the entire circuit, note that this current does not flow through each resistor. However, because

electric charge must be conserved in a circuit, the sum of the currents going through each branch of the circuit must add up to the current going through the battery. In other words, we cannot magically create charge somewhere in the circuit and add this new charge to the current. Let's check this reasoning by using Ohm's law to find the current through each resistor.

$$\begin{aligned} I_1 &= \frac{V}{R_1} = \frac{3\text{ V}}{10\ \Omega} = 0.30\text{ A} \\ I_2 &= \frac{V}{R_2} = \frac{3\text{ V}}{25\ \Omega} = 0.12\text{ A} \\ I_3 &= \frac{V}{R_3} = \frac{3\text{ V}}{15\ \Omega} = 0.20\text{ A} \end{aligned} \tag{19.32}$$

As expected, these currents add up to give 0.62 A, which is the total current found going through the equivalent resistor. Also, note that the smallest resistor has the largest current flowing through it, and vice versa.

WORKED EXAMPLE

Reasoning with Parallel Resistors

Without doing any calculation, what is the equivalent resistance of three identical resistors R in parallel?

STRATEGY

Three identical resistors R in parallel make three identical paths through which the current can flow. Thus, it is three times easier for the current to flow through these resistors than to flow through a single one of them.

Solution

If it is three times easier to flow through three identical resistors R than to flow through a single one of them, the equivalent resistance must be three times less: $R/3$.

Discussion

Let's check our reasoning by calculating the equivalent resistance of three identical resistors R in parallel. The equation for the equivalent resistance of resistors in parallel gives

$$\begin{aligned} R_{\text{equiv}} &= \frac{1}{1/R+1/R+1/R} \\ &= \frac{1}{3/R} \\ &= \frac{R}{3}. \end{aligned} \tag{19.33}$$

Thus, our reasoning was correct. In general, when more paths are available through which the current can flow, the equivalent resistance decreases. For example, if we have identical resistors R in parallel, the equivalent resistance would be $R/10$.

Practice Problems

10. Three resistors, 10, 20, and 30 Ω, are connected in parallel. What is the equivalent resistance?
    a. The equivalent resistance is 5.5 Ω
    b. The equivalent resistance is 60 Ω
    c. The equivalent resistance is 6 × 103 Ω
    d. The equivalent resistance is 6 × 104 Ω

11. If a 5-V drop occurs across R_1, and R_1 is connected in parallel to R_2, what is the voltage drop across R_2?
    a. Voltage drop across is 0 V.
    b. Voltage drop across is 2.5 V.
    c. Voltage drop across is 5 V.
    d. Voltage drop across is 10 V.

Resistors in Parallel and in Series

More complex connections of resistors are sometimes just combinations of series and parallel. Combinations of series and parallel resistors can be reduced to a single equivalent resistance by using the technique illustrated in Figure 19.17. Various parts are identified as either series or parallel, reduced to their equivalents, and further reduced until a single resistance is left. The

process is more time consuming than difficult.

Figure 19.17 This combination of seven resistors has both series and parallel parts. Each is identified and reduced to an equivalent resistance, and these are further reduced until a single equivalent resistance is reached.

Let's work through the four steps in Figure 19.17 to reduce the seven resistors to a single equivalent resistor. To avoid distracting algebra, we'll assume each resistor is 10 Ω . In step 1, we reduce the two sets of parallel resistors circled by the blue dashed loop. The upper set has three resistors in parallel and will be reduced to a single equivalent resistor R_{P1} . The lower set has two resistors in parallel and will be reduced to a single equivalent resistor R_{P2} . Using the equation for the equivalent resistance of resistors in parallel, we obtain

$$\begin{aligned} R_{P1} &= \frac{1}{1/R_2+1/R_3+1/R_4} = \frac{1}{1/10\ \Omega+1/10\ \Omega+1/10\ \Omega} = \frac{10}{3}\Omega \\ R_{P2} &= \frac{1}{1/R_5+1/R_6} = \frac{1}{1/10\ \Omega+1/10\ \Omega} = 5\ \Omega. \end{aligned} \tag{19.34}$$

These two equivalent resistances are encircled by the red dashed loop following step 1. They are in series, so we can use the

equation for the equivalent resistance of resistors in series to reduce them to a single equivalent resistance R_{S1} . This is done in step 2, with the result being

$$R_{S1} = R_{P1} + R_{P2} = \frac{10}{3}\Omega + 5\ \Omega = \frac{25}{3}\Omega. \tag{19.35}$$

The equivalent resistor R_{S1} appears in the green dashed loop following step 2. This resistor is in parallel with resistor R_7 , so the pair can be replaced by the equivalent resistor R_{P3} , which is given by

$$R_{P3} = \frac{1}{1/R_{S1} + 1/R_7} = \frac{1}{3/25\ \Omega + 1/10\ \Omega} = \frac{50}{11}\Omega. \tag{19.36}$$

This is done in step 3. The resistor R_{P3} is in series with the resistor R_1 , as shown in the purple dashed loop following step 3. These two resistors are combined in the final step to form the final equivalent resistor R_{equiv} , which is

$$R_{\text{equiv}} = R_1 + R_{P3} = 10\ \Omega + \frac{50}{11}\Omega = \frac{160}{11}\Omega. \tag{19.37}$$

Thus, the entire combination of seven resistors may be replaced by a single resistor with a resistance of about 14.5 Ω .

That was a lot of work, and you might be asking why we do it. It's important for us to know the equivalent resistance of the entire circuit so that we can calculate the current flowing through the circuit. Ohm's law tells us that the current flowing through a circuit depends on the resistance of the circuit and the voltage across the circuit. But to know the current, we must first know the equivalent resistance.

Here is a general approach to find the equivalent resistor for any arbitrary combination of resistors:

1. Identify a group of resistors that are only in parallel or only in series.
2. For resistors in series, use the equation for the equivalent resistance of resistors in series to reduce them to a single equivalent resistance. For resistors in parallel, use the equation for the equivalent resistance of resistors in parallel to reduce them to a single equivalent resistance.
3. Draw a new circuit diagram with the resistors from step 1 replaced by their equivalent resistor.
4. If more than one resistor remains in the circuit, return to step 1 and repeat. Otherwise, you are finished.

FUN IN PHYSICS

Robot

Robots have captured our collective imagination for over a century. Now, this dream of creating clever machines to do our dirty work, or sometimes just to keep us company, is becoming a reality. *Robotics* has become a huge field of research and development, with some technology already being commercialized. Think of the small autonomous vacuum cleaners, for example.

Figure 19.18 shows just a few of the multitude of different forms robots can take. The most advanced humanoid robots can walk, pour drinks, even dance (albeit not very gracefully). Other robots are bio-inspired, such as the *dogbot* shown in the middle photograph of Figure 19.18. This robot can carry hundreds of pounds of load over rough terrain. The photograph on the right in Figure 19.18 shows the inner workings of an *M-block,* developed by the Massachusetts Institute of Technology. These simple-looking blocks contain inertial wheels and electromagnets that allow them to spin and flip into the air and snap together in a variety of shapes. By communicating wirelessly between themselves, they self-assemble into a variety of shapes, such as desks, chairs, and someday maybe even buildings.

All robots involve an immense amount of physics and engineering. The simple act of pouring a drink has only recently been mastered by robots, after over 30 years of research and development! The balance and timing that we humans take for granted is in fact a very tricky act to follow, requiring excellent balance, dexterity, and feedback. To master this requires sensors to detect balance, computing power to analyze the data and communicate the appropriate compensating actions, and joints and actuators to implement the required actions.

In addition to sensing gravity or acceleration, robots can contain multiple different sensors to detect light, sound, temperature, smell, taste, etc. These devices are all based on the physical principles that you are studying in this text. For example, the optics used for robotic vision are similar to those used in your digital cameras: pixelated semiconducting detectors in which light is

converted into electrical signals. To detect temperature, simple thermistors may be used, which are resistors whose resistance changes depending on temperature.

Building a robot today is much less arduous than it was a few years ago. Numerous companies now offer kits for building robots. These range in complexity something suitable for elementary school children to something that would challenge the best professional engineers. If interested, you may find these easily on the Internet and start making your own robot today.

Figure 19.18 Robots come in many shapes and sizes, from the classic *humanoid* type to *dogbots* to small cubes that self-assemble to perform a variety of tasks.

WATCH PHYSICS

Resistors in Parallel

This video shows a lecturer discussing a simple circuit with a battery and a pair of resistors in parallel. He emphasizes that electrons flow in the direction opposite to that of the positive current and also makes use of the fact that the voltage is the same at all points on an ideal wire. The derivation is quite similar to what is done in this text, but the lecturer goes through it well, explaining each step.

Click to view content (https://www.openstax.org/l/28resistors)

GRASP CHECK

True or false—In a circuit diagram, we can assume that the voltage is the same at every point in a given wire.

a. false
b. true

WATCH PHYSICS

Resistors in Series and in Parallel

This video shows how to calculate the equivalent resistance of a circuit containing resistors in parallel and in series. The lecturer uses the same approach as outlined above for finding the equivalent resistance.

Click to view content (https://www.openstax.org/l/28resistorssp)

GRASP CHECK

Imagine connected N identical resistors in parallel. Each resistor has a resistance of R. What is the equivalent resistance for this group of parallel resistors?

a. The equivalent resistance is $(R)^N$.

b. The equivalent resistance is NR.
c. The equivalent resistance is $\frac{R}{N}$.
d. The equivalent resistance is $\frac{N}{R}$.

WORKED EXAMPLE

Find the Current through a Complex Resistor Circuit

The battery in the circuit below has a voltage rating of 10 V. What current flows through the circuit and in what direction?

STRATEGY

Apply the strategy for finding equivalent resistance to replace all the resistors with a single equivalent resistance, then use Ohm's law to find the current through the equivalent resistor.

Solution

The resistor combination R_4 and R_5 can be reduced to an equivalent resistance of

$$R_{P1} = \frac{1}{1/R_4 + 1/R_5} = \frac{1}{1/45\ \Omega + 1/60\ \Omega} = 25.71\ \Omega\text{R}. \tag{19.38}$$

Replacing R_4 and R_5 with this equivalent resistance gives the circuit below.

We now replace the two upper resistors R_2 and R_3 by the equivalent resistor R_{S1} and the two lower resistors R_{P1} and R_6 by their equivalent resistor R_{S2} . These resistors are in series, so we add them together to find the equivalent resistance.

$$\begin{aligned} R_{S1} &= R_2 + R_3 &= 50\ \Omega + 30\ \Omega &= 80\ \Omega \\ R_{S2} &= R_{P1} + R_6 &= 25.71\ \Omega + 20\ \Omega &= 45.71\ \Omega \end{aligned} \tag{19.39}$$

Replacing the relevant resistors with their equivalent resistor gives the circuit below.

Now replace the two resistors R_{S1} and R_{S2}, which are in parallel, with their equivalent resistor R_{P2}. The resistance of R_{P2} is

$$R_{P2} = \frac{1}{1/R_{S1} + 1/R_{S2}} = \frac{1}{1/80\ \Omega + 1/45.71\ \Omega} = 29.09\ \Omega. \tag{19.40}$$

Updating the circuit diagram by replacing R_{S1} and R_{S2} with this equivalent resistance gives the circuit below.

Finally, we combine resistors R_1 and R_{P2}, which are in series. The equivalent resistance is $R_{S3} = R_1 + R_{P2} = 75\ \Omega + 29.09\ \Omega = 104.09\ \Omega$. The final circuit is shown below.

We now use Ohm's law to find the current through the circuit.

$$\begin{aligned} V &= IR_{S3} \\ I &= \frac{V}{R_{S3}} = \frac{10\ \text{V}}{104.09\ \Omega} = 0.096\ \text{A} \end{aligned} \tag{19.41}$$

The current goes from the positive terminal of the battery to the negative terminal of the battery, so it flows clockwise in this circuit.

Discussion

This calculation may seem rather long, but with a little practice, you can combine some steps. Note also that extra significant digits were carried through the calculation. Only at the end was the final result rounded to two significant digits.

WORKED EXAMPLE

Strange-Looking Circuit Diagrams

Occasionally, you may encounter circuit diagrams that are not drawn very neatly, such as the diagram shown below. This circuit diagram looks more like how a real circuit might appear on the lab bench. What is the equivalent resistance for the resistors in this diagram, assuming each resistor is 10 Ω and the voltage rating of the battery is 12 V.

STRATEGY

Let's redraw this circuit diagram to make it clearer. Then we'll apply the strategy outlined above to calculate the equivalent resistance.

Solution

To redraw the diagram, consider the figure below. In the upper circuit, the blue resistors constitute a path from the positive terminal of the battery to the negative terminal. In parallel with this circuit are the red resistors, which constitute another path from the positive to negative terminal of the battery. The blue and red paths are shown more cleanly drawn in the lower circuit diagram. Note that, in both the upper and lower circuit diagrams, the blue and red paths connect the positive terminal of the battery to the negative terminal of the battery.

Now it is easier to see that R_1 and R_2 are in parallel, and the parallel combination is in series with R_4. This combination in turn is in parallel with the series combination of R_3 and R_5. First, we calculate the blue branch, which contains R_1, R_2, and R_4. The equivalent resistance is

$$R_{\text{blue}} = \frac{1}{1/R_1 + 1/R_2} + R_4 = \frac{1}{1/10\Omega + 1/10\ \Omega} + 10\ \Omega = 15\ \Omega. \tag{19.42}$$

where we show the contribution from the parallel combination of resistors and from the series combination of resistors. We now calculate the equivalent resistance of the red branch, which is

$$R_{\text{red}} = R_3 + R_5 = 10\ \Omega + 10\ \Omega = 20\ \Omega. \tag{19.43}$$

Inserting these equivalent resistors into the circuit gives the circuit below.

These two resistors are in parallel, so they can be replaced by a single equivalent resistor with a resistance of

$$R_{equiv} = \frac{1}{1/R_{blue} + 1/R_{red}} = \frac{1}{1/15\ \Omega + 1/20\ \Omega} = 8.6\ \Omega. \tag{19.44}$$

The final equivalent circuit is show below.

Discussion

Finding the equivalent resistance was easier with a clear circuit diagram. This is why we try to make clear circuit diagrams, where the resistors in parallel are lined up parallel to each other and at the same horizontal position on the diagram.

We can now use Ohm's law to find the current going through each branch to this circuit. Consider the circuit diagram with R_{blue} and R_{red} . The voltage across each of these branches is 12 V (i.e., the voltage rating of the battery). The current in the blue branch is

$$I_{blue} = \frac{V}{R_{blue}} = \frac{12\ \text{V}}{15\ \Omega} = 0.80\ \text{A}. \tag{19.45}$$

The current across the red branch is

$$I_{red} = \frac{V}{R_{red}} = \frac{12\ \text{V}}{20\ \Omega} = 0.60\ \text{A}. \tag{19.46}$$

The current going through the battery must be the sum of these two currents (can you see why?), or 1.4 A.

Practice Problems

12. What is the formula for the equivalent resistance of two parallel resistors with resistance R_1 and R_2?

a. Equivalent resistance of two parallel resistors $R_{eqv} = R_1 + R_2$
b. Equivalent resistance of two parallel resistors $R_{eqv} = R_1 \times R_2$
c. Equivalent resistance of two parallel resistors $R_{eqv} = R_1 - R_2$
d. Equivalent resistance of two parallel resistors $R_{eqv} = \frac{1}{1/R_1 + 1/R_2}$

13.

Figure 19.19

What is the equivalent resistance for the two resistors shown below?

a. The equivalent resistance is 20 Ω
b. The equivalent resistance is 21 Ω
c. The equivalent resistance is 90 Ω
d. The equivalent resistance is 1,925 Ω

Check Your Understanding

14. The voltage drop across parallel resistors is _____.

a. the same for all resistors
b. greater for the larger resistors
c. less for the larger resistors
d. greater for the smaller resistors

15. Consider a circuit of parallel resistors. The smallest resistor is 25 Ω . What is the upper limit of the equivalent resistance?

a. The upper limit of the equivalent resistance is 2.5 Ω.
b. The upper limit of the equivalent resistance is 25 Ω.
c. The upper limit of the equivalent resistance is 100 Ω.
d. There is no upper limit.

19.4 Electric Power

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Define electric power and describe the electric power equation
- Calculate electric power in circuits of resistors in series, parallel, and complex arrangements

Section Key Terms

electric power

Power is associated by many people with electricity. Every day, we use electric power to run our modern appliances. Electric power transmission lines are visible examples of electricity providing power. We also use electric power to start our cars, to run our computers, or to light our homes. Power is the rate at which energy of any type is transferred; **electric power** is the rate at which electric energy is transferred in a circuit. In this section, we'll learn not only what this means, but also what factors determine electric power.

To get started, let's think of light bulbs, which are often characterized in terms of their power ratings in watts. Let us compare a 25-W bulb with a 60-W bulb (see Figure 19.20). Although both operate at the same voltage, the 60-W bulb emits more light intensity than the 25-W bulb. This tells us that something other than voltage determines the power output of an electric circuit.

Incandescent light bulbs, such as the two shown in Figure 19.20, are essentially resistors that heat up when current flows through them and they get so hot that they emit visible and invisible light. Thus the two light bulbs in the photo can be considered as two different resistors. In a simple circuit such as a light bulb with a voltage applied to it, the resistance determines the current by Ohm's law, so we can see that current as well as voltage must determine the power.

Figure 19.20 On the left is a 25-W light bulb, and on the right is a 60-W light bulb. Why are their power outputs different despite their operating on the same voltage?

The formula for power may be found by dimensional analysis. Consider the units of power. In the SI system, power is given in watts (W), which is energy per unit time, or J/s

$$\mathrm{W} = \frac{\mathrm{J}}{\mathrm{s}}. \tag{19.47}$$

Recall now that a voltage is the potential energy per unit charge, which means that voltage has units of J/C

$$\mathrm{V} = \frac{\mathrm{J}}{\mathrm{C}}. \tag{19.48}$$

We can rewrite this equation as $\mathrm{J} = \mathrm{V} \times \mathrm{C}$ and substitute this into the equation for watts to get

$$\mathrm{W} = \frac{\mathrm{J}}{\mathrm{s}} = \frac{\mathrm{V} \times \mathrm{C}}{\mathrm{s}} = \mathrm{V} \times \frac{\mathrm{C}}{\mathrm{s}}.$$

But a Coulomb per second (C/s) is an electric current, which we can see from the definition of electric current, $I = \frac{\Delta Q}{\Delta t}$, where ΔQ is the charge in coulombs and Δt is time in seconds. Thus, equation above tells us that electric power is voltage times current, or

$$P = IV.$$

This equation gives the electric power consumed by a circuit with a voltage drop of V and a current of I.

For example, consider the circuit in Figure 19.21. From Ohm's law, the current running through the circuit is

$$I = \frac{V}{R} = \frac{12\ \mathrm{V}}{100\ \Omega} = 0.12\ \mathrm{A}. \tag{19.49}$$

Thus, the power consumed by the circuit is

$$P = VI = (12\ \mathrm{V})(0.12\ \mathrm{A}) = 1.4\ \mathrm{W}. \tag{19.50}$$

Where does this power go? In this circuit, the power goes primarily into heating the resistor in this circuit.

Figure 19.21 A simple circuit that consumes electric power.

In calculating the power in the circuit of Figure 19.21, we used the resistance and Ohm's law to find the current. Ohm's law gives the current: $I = V/R$, which we can insert into the equation for electric power to obtain

$$P = IV = \left(\frac{V}{R}\right) V = \frac{V^2}{R}.$$

This gives the power in terms of only the voltage and the resistance.

We can also use Ohm's law to eliminate the voltage in the equation for electric power and obtain an expression for power in terms of just the current and the resistance. If we write Ohm's law as $V = IR$ and use this to eliminate V in the equation $P = IV$, we obtain

$$P = IV = I(IR) = I^2R.$$

This gives the power in terms of only the current and the resistance.

Thus, by combining Ohm's law with the equation $P = IV$ for electric power, we obtain two more expressions for power: one in terms of voltage and resistance and one in terms of current and resistance. Note that only resistance (not capacitance or anything else), current, and voltage enter into the expressions for electric power. This means that the physical characteristic of a circuit that determines how much power it dissipates is its resistance. Any capacitors in the circuit do not dissipate electric power—on the contrary, capacitors either store electric energy or release electric energy back to the circuit.

To clarify how voltage, resistance, current, and power are all related, consider Figure 19.22, which shows the *formula wheel*. The quantities in the center quarter circle are equal to the quantities in the corresponding outer quarter circle. For example, to express a potential V in terms of power and current, we see from the formula wheel that $V = P/I$.

Figure 19.22 The formula wheel shows how volts, resistance, current, and power are related. The quantities in the inner quarter circles equal the quantities in the corresponding outer quarter circles.

WORKED EXAMPLE

Find the Resistance of a Lightbulb

A typical older incandescent lightbulb was 60 W. Assuming that 120 V is applied across the lightbulb, what is the current through the lightbulb?

STRATEGY

We are given the voltage and the power output of a simple circuit containing a lightbulb, so we can use the equation $P = IV$ to find the current I that flows through the lightbulb.

Solution

Solving $P = IV$ for the current and inserting the given values for voltage and power gives

$$\begin{aligned} P &= IV \\ I &= \frac{P}{V} = \frac{60\text{ W}}{120\text{ V}} = 0.50\text{ A}. \end{aligned} \tag{19.51}$$

Thus, a half ampere flows through the lightbulb when 120 V is applied across it.

Discussion

This is a significant current. Recall that household power is AC and not DC, so the 120 V supplied by household sockets is an alternating power, not a constant power. The 120 V is actually the time-averaged power provided by such sockets. Thus, the average current going through the light bulb over a period of time longer than a few seconds is 0.50 A.

WORKED EXAMPLE

Boot Warmers

To warm your boots on cold days, you decide to sew a circuit with some resistors into the insole of your boots. You want 10 W of heat output from the resistors in each insole, and you want to run them from two 9-V batteries (connected in series). What total resistance should you put in each insole?

STRATEGY

We know the desired power and the voltage (18 V, because we have two 9-V batteries connected in series), so we can use the equation $P = V^2/R$ to find the requisite resistance.

Solution

Solving $P = V^2/R$ for the resistance and inserting the given voltage and power, we obtain

$$\begin{aligned} P &= \frac{V^2}{R} \\ R &= \frac{V^2}{P} = \frac{(18\ \text{V})^2}{10\ \text{W}} = 32\ \Omega. \end{aligned} \tag{19.52}$$

Thus, the total resistance in each insole should be 32 Ω.

Discussion

Let's see how much current would run through this circuit. We have 18 V applied across a resistance of 32 Ω, so Ohm's law gives

$$I = \frac{V}{R} = \frac{18\ \text{V}}{32\ \Omega} = 0.56\ \text{A}. \tag{19.53}$$

All batteries have labels that say how much charge they can deliver (in terms of a current multiplied by a time). A typical 9-V alkaline battery can deliver a charge of 565 $\text{mA} \cdot \text{h}$ (so two 9 V batteries deliver 1,130 $\text{mA} \cdot \text{h}$), so this heating system would function for a time of

$$t = \frac{1130 \times 10^{-3}\ \text{A} \cdot \text{h}}{0.56\ \text{A}} = 2.0\ \text{h}. \tag{19.54}$$

WORKED EXAMPLE

Power through a Branch of a Circuit

Each resistor in the circuit below is 30 Ω . What power is dissipated by the middle branch of the circuit?

STRATEGY

The middle branch of the circuit contains resistors R_3 and R_5 in series. The voltage across this branch is 12 V. We will first find the equivalent resistance in this branch, and then use $P = V^2/R$ to find the power dissipated in the branch.

Solution

The equivalent resistance is $R_{\text{middle}} = R_3 + R_5 = 30\ \Omega + 30\ \Omega = 60\ \Omega$. The power dissipated by the middle branch of the circuit is

$$P_{\text{middle}} = \frac{V^2}{R_{\text{middle}}} = \frac{(12\text{ V})^2}{60\ \Omega} = 2.4\text{ W}. \tag{19.55}$$

Discussion

Let's see if energy is conserved in this circuit by comparing the power dissipated in the circuit to the power supplied by the battery. First, the equivalent resistance of the left branch is

$$R_{\text{left}} = \frac{1}{1/R_1 + 1/R_2} + R_4 = \frac{1}{1/30\ \Omega + 1/30\ \Omega} + 30\ \Omega = 45\ \Omega. \tag{19.56}$$

The power through the left branch is

$$P_{\text{left}} = \frac{V^2}{R_{\text{left}}} = \frac{(12\text{ V})^2}{45\ \Omega} = 3.2\text{ W}. \tag{19.57}$$

The right branch contains only R_6, so the equivalent resistance is $R_{\text{right}} = R_6 = 30\ \Omega$. The power through the right branch is

$$P_{\text{right}} = \frac{V^2}{R_{\text{right}}} = \frac{(12\text{ V})^2}{30\ \Omega} = 4.8\text{ W}. \tag{19.58}$$

The total power dissipated by the circuit is the sum of the powers dissipated in each branch.

$$P = P_{\text{left}} + P_{\text{middle}} + P_{\text{right}} = 2.4\text{ W} + 3.2\text{ W} + 4.8\text{ W} = 10.4\text{ W} \tag{19.59}$$

The power provided by the battery is

$$P = IV. \tag{19.60}$$

where I is the total current flowing through the battery. We must therefore add up the currents going through each branch to obtain I. The branches contributes currents of

$$\begin{aligned} I_{\text{left}} &= \frac{V}{R_{\text{left}}} = \frac{12\text{ V}}{45\ \Omega} = 0.2667\text{ A} \\ \mathrm{I}_{\text{middle}} &= \frac{V}{R_{\text{middle}}} = \frac{12\text{ V}}{60\ \Omega} = 0.20\text{ A} \\ \mathrm{I}_{\text{right}} &= \frac{V}{R_{\text{right}}} = \frac{12\text{ V}}{30\ \Omega} = 0.40\text{ A}. \end{aligned} \tag{19.61}$$

The total current is

$$I = I_{\text{left}} + I_{\text{middle}} + I_{\text{right}} = 0.2667\text{ A} + 0.20\text{ A} + 0.40\text{ A} = 0.87\text{ A}. \tag{19.62}$$

and the power provided by the battery is

$$P = IV = (0.87\text{ A})(12\text{ V}) = 10.4\text{ W}. \tag{19.63}$$

This is the same power as is dissipated in the resistors of the circuit, which shows that energy is conserved in this circuit.

Practice Problems

16. What is the formula for the power dissipated in a resistor?

a. The formula for the power dissipated in a resistor is $P = \frac{I}{V}$.
b. The formula for the power dissipated in a resistor is $P = \frac{V}{I}$.
c. The formula for the power dissipated in a resistor is $P = IV$.
d. The formula for the power dissipated in a resistor is $P = I^2V$.

17. What is the formula for power dissipated by a resistor given its resistance and the voltage across it?

a. The formula for the power dissipated in a resistor is $P = \frac{R}{V^2}$
b. The formula for the power dissipated in a resistor is $P = V^2R$
c. The formula for the power dissipated in a resistor is $P = \frac{V^2}{R}$
d. The formula for the power dissipated in a resistor is $P = I^2R$

Check your Understanding

18. Which circuit elements dissipate power?

a. capacitors
b. inductors
c. ideal switches
d. resistors

19. Explain in words the equation for power dissipated by a given resistance.

a. Electric power is proportional to current through the resistor multiplied by the square of the voltage across the resistor.
b. Electric power is proportional to square of current through the resistor multiplied by the voltage across the resistor.
c. Electric power is proportional to current through the resistor divided by the voltage across the resistor.
d. Electric power is proportional to current through the resistor multiplied by the voltage across the resistor.

KEY TERMS

alternating current electric current whose direction alternates back and forth at regular intervals

ampere unit for electric current; one ampere is one coulomb per second ($1\ \text{A} = 1\ \text{C/s}$)

circuit diagram schematic drawing of an electrical circuit including all circuit elements, such as resistors, capacitors, batteries, and so on

conventional current flows in the direction that a positive charge would flow if it could move

direct current electric current that flows in a single direction

electric circuit physical network of paths through which electric current can flow

electric current electric charge that is moving

electric power rate at which electric energy is transferred in a circuit

equivalent resistor resistance of a single resistor that is the same as the combined resistance of a group of resistors

in parallel when a group of resistors are connected side by side, with the top ends of the resistors connected together by a wire and the bottom ends connected together by a different wire

in series when elements in a circuit are connected one after the other in the same branch of the circuit

nonohmic material that does not follow Ohm's law

Ohm's law electric current is proportional to the voltage applied across a circuit or other path

ohmic material that obeys Ohm's law

resistance how much a circuit element opposes the passage of electric current; it appears as the constant of proportionality in Ohm's law

resistor circuit element that provides a known resistance

steady state when the characteristics of a system do not change over time

SECTION SUMMARY

19.1 Ohm's law

- Direct current is constant over time; alternating current alternates smoothly back and forth over time.
- Electrical resistance causes materials to extract work from the current that flows through them.
- In ohmic materials, voltage drop along a path is proportional to the current that runs through the path.

19.2 Series Circuits

- Circuit diagrams are schematic representations of electric circuits.
- Resistors in series are resistors that are connected head to tail.
- The same current runs through all resistors in series; however, the voltage drop across each resistor can be different.
- The voltage is the same at every point in a given wire.

19.3 Parallel Circuits

- The equivalent resistance of a group of N identical resistors R connected in parallel is R/N.
- Connecting resistors in parallel provides more paths for the current to go through, so the equivalent resistance is always less than the smallest resistance of the parallel resistors.
- The same voltage drop occurs across all resistors in parallel; however, the current through each resistor can differ.

19.4 Electric Power

- Electric power is dissipated in the resistances of a circuit. Capacitors do not dissipate electric power.
- Electric power is proportional to the voltage and the current in a circuit.
- Ohm's law provides two extra expressions for electric power: one that does not involve current and one that does not involve voltage.

KEY EQUATIONS

19.1 Ohm's law

electric current I is the charge ΔQ that passes a plane per unit time Δt	$I = \frac{\Delta Q}{\Delta t}$
an ampere is the coulombs per unit time that pass a plane	$1\ \text{A} = 1\ \text{C/s}$
Ohm's law: the current I is proportional to the voltage V, with the resistance R being the constant of proportionality	$V = IR$

19.2 Series Circuits

the equivalent resistance of N resistors connected in series	$R_{\text{equiv}} = R_1 + R_2 + \cdots + R_N$

19.3 Parallel Circuits

the equivalent resistance of N resistors connected in parallel	$R_{\text{equiv}} = \frac{1}{1/R_1 + 1/R_2 + \cdots + 1/R_N}$

19.4 Electric Power

for a given current I flowing through a potential difference V, the electric power dissipated	$P = IV$
for a given current I flowing through a resistance R, the electric power dissipated	$P = I^2R$
for a given voltage difference V across a resistor R, the electric power dissipated	$P = \frac{V^2}{R}$

CHAPTER REVIEW

Concept Items

19.1 Ohm's law

1. You connect a resistor across a battery. In which direction do the electrons flow?
   a. The electrons flow from the negative terminal of the battery to the positive terminal of the battery.
   b. The electrons flow from the positive terminal of the battery to the negative terminal of the battery.

2. How does current depend on resistance in Ohm's law?
   a. Current is directly proportional to the resistance.
   b. Current is inversely proportional to the resistance.
   c. Current is proportional to the square of the resistance.
   d. Current is inversely proportional to the square of the resistance.

3. In the context of electricity, what is resistance?
   a. Resistance is the property of materials to resist the passage of voltage.
   b. Resistance is the property of materials to resist the passage of electric current.
   c. Resistance is the property of materials to increase the passage of voltage.
   d. Resistance is the property of materials to increase the passage of electric current.

4. What is the mathematical formula for Ohm's law?
   a. $V = I^2R$
   b. $V = \frac{R}{I}$
   c. $V = \frac{I}{R}$
   d. $V = IR$

19.2 Series Circuits

5. In which circuit are all the resistors connected in series?

a.

b.

c.

d.

6. What is the voltage and current through the capacitor in the circuit below a long time after the switch is closed?

   a. 0 V, 0 A
   b. 0 V, 10 A
   c. 10 V, 0 A
   d. 10 V, 10 A

19.3 Parallel Circuits

7. If you remove resistance from a circuit, does the total resistance of the circuit always decrease? Explain.
   a. No, because for parallel combination of resistors, the resistance through the remaining circuit increases.
   b. Yes, because for parallel combination of resistors, the resistance through the remaining circuit increases.

8. Explain why the equivalent resistance of a parallel combination of resistors is always less than the smallest of the parallel resistors.
   a. Adding resistors in parallel gives the current a shorter path through which it can flow hence decreases the overall resistance.
   b. Adding resistors in parallel gives the current another path through which it can flow hence decreases the overall resistance.
   c. Adding resistors in parallel reduce the number of paths through which the current can flow hence decreases the overall resistance.
   d. Adding resistors in parallel gives the current longer path through which it can flow hence decreases the overall resistance.

19.4 Electric Power

9. To draw the most power from a battery, should you connect a small or a large resistance across its terminals? Explain.
   a. Small resistance, because smaller resistance will lead to the largest power
   b. Large resistance, because smaller resistance will lead to the largest power

10. If you double the current through a resistor, by what factor does the power dissipated by the resistor change?
   a. Power increases by a factor of two.
   b. Power increases by a factor of four.
   c. Power increases by a factor of eight.
   d. Power increases by a factor of 16.

Critical Thinking Items

19.1 Ohm's law

11. An accelerator accelerates He nuclei (change = $2e$) to a speed of $v = 2 \times 10^6$ m/s. What is the current if the linear density of He nuclei is $\lambda = 108\ \text{m}^{-1}$?
   a. $I = 9.6 \times 10^{-5}$ A
   b. $I = 3.2 \times 10^{-5}$ A
   c. $I = 12.8 \times 10^{-5}$ A
   d. $I = 6.4 \times 10^{-5}$ A

12. How can you verify whether a certain material is ohmic?
   a. Make a resistor from this material and measure the current going through this resistor for several different voltages. If the current is proportional to the voltage, then the material is ohmic.
   b. Make a resistor from this material and measure the current going through this resistor for several different voltages. If the current is inversely proportional to the voltage, then the material is ohmic.
   c. Make a resistor from this material and measure the current going through this resistor for several different voltages. If the current is proportional to the square of the voltage, then the material is ohmic.
   d. Make a resistor from this material and measure the current going through this resistor for several different voltages. If the current is inversely proportional to the square of the voltage, then the material is ohmic.

19.2 Series Circuits

13. Given three batteries (5V, 9V, 12V) and five resistors (10, 20, 30, 40, 50Ω) to choose from, what can you choose to form a circuit diagram with a current of 0.175A? You do not need to use all of the components.
   a. Batteries (5V, 9V) and resistors (30Ω, 50Ω) connected in series
   b. Batteries (5V4, 12V) and resistors (10Ω, 20Ω, 40Ω, and 50Ω) connected in series.
   c. Batteries (5V, 9V, and 12V) and resistors (10Ω, 20Ω, and 30Ω) connected in series.

14. What is the maximum resistance possible given a resistor of 100 and a resistor of 40 Ω?
   a. 100 Ω
   b. 140 Ω
   c. 180 Ω
   d. 240 Ω

15. Rank the points A, B, C, and D in the circuit diagram from lowest voltage to highest voltage.

a. A, B, C, D
b. B, C, A, D
c. C, B, A, D
d. D, A, B, C

19.3 Parallel Circuits

16. Can all resistor combinations be reduced to series and parallel combinations?
   a. No, all practical resistor circuits cannot be reduced to series and parallel combinations.
   b. Yes, all practical resistor circuits can be reduced to series and parallel combinations.

17. What is the equivalent resistance of the circuit shown below?

Figure 19.23

   a. The equivalent resistance of the circuit 14 Ω.
   b. The equivalent resistance of the circuit 16.7 Ω.
   c. The equivalent resistance of the circuit 140 Ω.
   d. The equivalent resistance of the circuit 195 Ω.

19.4 Electric Power

18. Two lamps have different resistances. (a) If the lamps are connected in parallel, which one is brighter, the lamp with greater resistance or the lamp with less resistance? (b) If the lamps are connected in series, which one is brighter? Note that the brighter lamp dissipates more power.
   a. (a) lamp with greater resistance; (b) lamp with less resistance
   b. (a) lamp with greater resistance; (b) lamp with greater resistance
   c. (a) lamp with less resistance; (b) lamp with less resistance
   d. (a) lamp with less resistance; (b) lamp with greater resistance

19. To measure the power consumed by your laptop computer, you place an ammeter (a device that measures electric current) in series with its DC power supply. When the screen is off, the computer draws 0.40 A of current.
When the screen is on at full brightness, it draws 0.90 A of current. Knowing the DC power supply delivers 16 V, how much power is used by the screen?
   a. The power used by the screen is −8.0 W.
   b. The power used by the screen is 0.3 W.
   c. The power used by the screen is 3.2 W.
   d. The power used by the screen is 8.0 W.

Problems

19.1 Ohm's law

20. What voltage is needed to make 6 C of charge traverse a 100-Ω resistor in 1 min?
   a. The required voltage is 1×10^{-3} V.
   b. The required voltage is 10 V.
   c. The required voltage is 1,000 V.
   d. The required voltage is 10,000 V.

21. Resistors typically obey Ohm's law at low currents, but show deviations at higher currents because of heating. Suppose you were to conduct an experiment measuring the voltage, *V*, across a resistor as a function of current, *I*, including currents whose deviations from Ohm's law start to become apparent. For a data plot of *V* versus *I*, which of the following functions would be best to fit the data? Assume that *a*, *b*, and *c* are nonzero constants adjusted to fit the data.
   a. $V = aI$
   b. $V = aI + b$
   c. $V = aI + bI^2$
   d. $V = aI + bI^2 + c$

22. A battery of unknown voltage V_1 is attached across a resistor R_1. You add a second battery with $V_2 = 9.0$ V in series with V_1 so that the voltage across R_1 is now $V_1 + V_2$ and measure 0.3 A of current through resistor R_1. You add a third battery with $V_3 = 9.0$ V in series with the first two batteries so that the voltage across R_3 is $V_1 + V_2 + V_3$ and measure 0.4 A of current through R_1. What is the resistance of R_1?

a. 23.25 Ω
b. 21.75 Ω
c. 31.33 Ω
d. 13.0 Ω

19.2 Series Circuits

23. What is the voltage drop across two 80-Ω resistors connected in series with 0.15 A flowing through them?
a. 12 V
b. 24 V
c. 36 V
d. 48 V

24. In this circuit, the voltage drop across the upper resistor is 4.5 V. What is the battery voltage?

a. 4.5V
b. 7.5V
c. 12V
d. 18V

19.3 Parallel Circuits

25. What is the equivalent resistance of this circuit?

a. The equivalent resistance of the circuit is 32.7 Ω.
b. The equivalent resistance of the circuit is 100 Ω.
c. The equivalent resistance of the circuit is 327 Ω.
d. The equivalent resistance of the circuit is 450 Ω.

26. What is the equivalent resistance of the circuit shown below?

a. The equivalent resistance is 25 Ω.
b. The equivalent resistance is 50 Ω.
c. The equivalent resistance is 75 Ω.
d. The equivalent resistance is 100 Ω.

19.4 Electric Power

27. When 12 V are applied across a resistor, it dissipates 120 W of power. What is the current through the resistor?
a. The current is 1,440 A.
b. The current is 10 A.
c. The current is 0.1 A.
d. The current is 0.01 A.

28. Warming 1 g of water requires 1 J of energy per . How long would it take to warm 1 L of water from 20 to 40 °C if you immerse in the water a 1-kW resistor connected across a 9.0-V batteries aligned in series?
a. 10 min
b. 20 min
c. 30 min
d. 40 min

Performance Task

19.4 Electric Power

29.
1. An incandescent light bulb (i.e., an old-fashioned light bulb with a little wire in it)
2. A lightbulb socket to hold the light bulb
3. A variable voltage source
4. An ammeter

Procedure

- Screw the lightbulb into its socket. Connect the positive terminal of the voltage source to the input of the ammeter. Connect the output of the ammeter to one connection of the socket. Connect the other connection of the socket to the negative terminal of the voltage source. Ensure that the voltage source is set to supply DC voltage and that the ammeter is set to measure DC amperes. The desired circuit is shown below.

- On a piece of paper, make a two-column table with 10 rows. Label the left column *volts* and the right column *current*. Adjust the voltage source so that it supplies from between 1 and 10 volts DC. For each voltage, write the voltage in the volts column and the corresponding amperage measured by the ammeter in the current column. Make a plot of volts versus current, that is, a plot with volts on the vertical axis and current on the horizontal axis. Use this data and the plot to answer the following questions:

1. What is the resistance of the lightbulb?
2. What is the range of possible error in your result for the resistance?
3. In a single word, how would you describe the curve formed by the data points?

TEST PREP

Multiple Choice

19.1 Ohm's law

30. What are the SI units for electric current?
- a. C/s
- b. e/s
- c. −e/s
- d. C/s^2

31. What is the SI unit for resistance?
- a. C/m
- b. C/s
- c. Ω
- d. Ψ

32. The equivalent unit for an ohm is a _____.
- a. V/A
- b. C/m
- c. $\frac{A}{V}$
- d. V/s

33. You put 9.0 V DC across resistor R_1 and measure the current through it. With the same voltage across resistor R_2, you measure twice the current. What is the ratio $\frac{R_1}{R_2}$?
- a. 1
- b. $\frac{1}{2}$
- c. 4
- d. 2

19.2 Series Circuits

34. What does the circuit element shown represent?

- a. a battery
- b. a capacitor
- c. the ground
- d. a switch

35. How many 10-Ω resistors must be connected in series to make an equivalent resistance of 80 Ω?
- a. 80
- b. 8
- c. 20
- d. 40

36. Which two circuit elements are represented in the circuit diagram?

- a. a battery connected in series with an inductor
- b. a capacitor connected in series with a resistor
- c. a resistor connected in series with a battery
- d. an inductor connected in series with a resistor

37. How much current will flow through a 10-V battery with a 100-Ω resistor connected across its terminals?
- a. 0.1 A
- b. 1.0 A
- c. 0
- d. 1,000 A

19.3 Parallel Circuits

38. A 10-Ω resistor is connected in parallel to another resistor R . The equivalent resistance of the pair is 8 Ω. What is the resistance R?
- a. 10 Ω
- b. 20 Ω
- c. 30 Ω

d. 40 Ω

39. Are the resistors shown connected in parallel or in series? Explain.

a. The resistors are connected in parallel because the same current flows through all three resistors.
b. The resistors are connected in parallel because different current flows through all three resistors.
c. The resistors are connected in series because the same current flows through all three resistors.
d. The resistors are connected in series because different current flows through all three resistors.

19.4 Electric Power

40. Which equation below for electric power is incorrect?
a. $P = I^2R$
b. $P = \frac{V}{R^2}$
c. $P = IV$
d. $P = \frac{V^2}{R}$

41. What power is dissipated in a circuit through which 0.12 A flows across a potential drop of 3.0 V?
a. 0.36 W
b. 0.011 W
c. Voltage drop across is 5 V.
d. 2.5 W

42. How does a resistor dissipate power?
a. A resistor dissipates power in the form of heat.
b. A resistor dissipates power in the form of sound.
c. A resistor dissipates power in the form of light.
d. A resistor dissipates power in the form of charge.

43. What power is dissipated in a circuit through which 0.12 A flows across a potential drop of 3.0 V?
a. 0.36 W
b. 0.011 W
c. 5 V
d. 2.5 W

Short Answer

19.1 Ohm's law

44. True or false—it is possible to produce nonzero DC current by adding together AC currents.
a. false
b. true

45. What type of current is used in cars?
a. alternating current
b. indirect current
c. direct current
d. straight current

46. If current were represented by C, voltage by B, and resistance by g, what would the mathematical expression be for Ohm's law?
a. $C = Bg$
b. $g = BC$
c. $\frac{B}{C} = \frac{C}{g}$
d. $B = Cg$

47. Give a verbal expression for Ohm's law.
a. Ohm's law says that the current through a resistor equals the voltage across the resistor multiplied by the resistance of the resistor.
b. Ohm's law says that the voltage across a resistor equals the current through the resistor multiplied by the resistance of the resistor.
c. Ohm's law says that the resistance of the resistor equals the current through the resistor multiplied by the voltage across a resistor.
d. Ohm's law says that the voltage across a resistor equals the square of the current through the resistor multiplied by the resistance of the resistor.

48. What is the current through a 100-Ω resistor with 12 V across it?
a. 0
b. 0.12 A
c. 8.33 A
d. 1,200 A

49. What resistance is required to produce 0.15 A from a 9.0 V battery?
a. 0.017
b. 1
c. 60
d. 120

19.2 Series Circuits

50. Given a circuit with one 9-V battery and with its negative terminal connected to ground. The two paths are connected to ground from the positive terminal: the right path with a 20-Ω and a 100-Ω resistor and the left path with a 50-Ω resistor. How much current will flow in the right branch?
a. $\frac{9}{120}$
b. $\frac{9}{100}$

c. $\frac{9}{50}$
d. $\frac{9}{20}$

51. Through which branch in the circuit below does the most current flow?

a. All of the current flows through the left branch due to the open switch.
b. All of the current flows through the right branch due to the open switch in the left branch.
c. All of the current flows through the middle branch due to the open switch in the left branch
d. There will be no current in any branch of the circuit due to the open switch.

52. What current flows through the 75-Ω resistor in the circuit below?

a. 0.072 A
b. 0.12 A
c. 0.18 A
d. 0.3 A

53. What is the equivalent resistance for the circuit below if $V = 9.0$ V and $I = 0.25$ A?

a. 0.028 Ω
b. 2.25 Ω
c. 36 Ω
d. 72 Ω

19.3 Parallel Circuits

54. Ten 100-Ω resistors are connected in series. How can you increase the total resistance of the circuit by about 40 percent?
a. Adding two 10-Ω resistors increases the total resistance of the circuit by about 40 percent.
b. Removing two 10-Ω resistors increases the total resistance of the circuit by about 40 percent.
c. Adding four 10-Ω resistors increases the total resistance of the circuit by about 40 percent.
d. Removing four 10-Ω resistors increases the total resistance of the circuit by about 40 percent.

55. Two identical resistors are connected in parallel across the terminals of a battery. If you increase the resistance of one of the resistors, what happens to the current through and the voltage across the other resistor?
a. The current and the voltage remain the same.
b. The current decreases and the voltage remains the same.
c. The current and the voltage increases.
d. The current increases and the voltage remains the same.

56.

In the circuit below, through which resistor(s) does the most current flow? Through which does the least flow? Explain.
a. The most current flows through the 15-Ω resistor because all the current must pass through this resistor.
b. The most current flows through the 20-Ω resistor because all the current must pass through this resistor.
c. The most current flows through the 25-Ω resistor because it is the highest resistance.
d. The same current flows through the all the resistor because all the current must pass through each of the resistors.

19.4 Electric Power

57. You want to increase the power dissipated in a circuit.

You have the choice between doubling the current or doubling the resistance, with the voltage remaining constant. Which one would you choose?

a. doubling the resistance
b. doubling the current

58. You want to increase the power dissipated in a circuit. You have the choice between reducing the voltage or reducing the resistance, with the current remaining constant. Which one would you choose?

a. reduce the voltage to increase the power
b. reduce the resistance to increase the power

59. What power is dissipated in the circuit consisting of 310-Ω resistors connected in series across a 9.0-V battery?

a. The power dissipated is 2430 W.
b. The power dissipated is 270 W.
c. The power dissipated is 2.7 W.
d. The power dissipated is 0.37 W.

60. What power is dissipated in a circuit consisting of three 10-Ω resistors connected in parallel across a 9.0-V battery?

a. The power dissipated is 270 W.
b. The power dissipated is 30 W.
c. The power dissipated is 24 W.
d. The power dissipated is 1/24 W.

Extended Response

19.1 Ohm's law

61. Describe the relationship between current and charge. Include an explanation of how the direction of the current is defined.

a. Electric current is the charge that passes through a conductor per unit time. The direction of the current is defined to be the direction in which positive charge would flow.
b. Electric current is the charges that move in a conductor. The direction of the current is defined to be the direction in which positive charge would flow.
c. Electric current is the charge that passes through a conductor per unit time. The direction of the current is defined to be the direction in which negative charge would flow.
d. Electric current is the charges that move in a conductor. The direction of the current is defined to be the direction in which negative charge would flow.

62. What could cause Ohm's law to break down?

a. If small amount of current flows through a resistor, the resistor will heat up so much that it will change state, in violation of Ohm's law.
b. If excessive amount of current flows through a resistor, the resistor will heat up so much that it will change state, in violation of Ohm's law.
c. If small amount of current flows through a resistor, the resistor will not heat up so much and it will not change its state, in violation of Ohm's law.
d. If excessive amount of current flows through a resistor, the resistor will heat up so much that it will not change its state, in violation of Ohm's law.

63. You connect a single resistor R across a 10-V battery and find that 0.01 A flows through the circuit. You add another resistor R after the first resistor and find that 0.005 A flows through the circuit. If you have 10 resistors R connected in a line one after the other, what would be their total resistance?

a. $\frac{R}{10}$
b. 5 R
c. $\frac{10}{R}$
d. 10 R

19.2 Series Circuits

64. Explain why the current is the same at all points in the circuit below.

a. If the current were not constant, the mobile charges would bunch up in places, which means that the voltage would decrease at that point. A lower voltage at some point would push the current in the direction that further decreases the voltage.
b. If the current were not constant, the mobile charges would bunch up in places, which means that the voltage would increase at that point. But a higher voltage at some point would push the current in the direction that decreases the voltage.
c. If the current were not constant, the mobile charges would bunch up in places, which mean that the voltage would increase at that point. A higher voltage at some point would push the current in the direction that further increases the

voltage.

d. If the current were not constant, the mobile charges would bunch up in places, which mean that the voltage would decrease at that point. But a lower voltage at some point would push the current in the direction that increases the voltage.

65. What is the current through each resistor in the circuit?

a. Current through resistors R_1, R_2, R_3, and R_4 is 0.48 A, 0.30 A, 1.2 A, and 0.24 A, respectively.

b. Current through resistors R_1, R_2, R_3, and R_4 is 1200 A, 1920 A, 480 A, and 2400 A, respectively.

c. Current through resistors R_1, R_2, R_3, and is R_4 2.08 A, 3.34 A, 0.833 A, and 4.17 A, respectively.

d. The same amount of current, 0.096 A, flows through all of the resistors.

19.3 Parallel Circuits

66. In a house, a single incoming wire at a high potential with respect to the ground provides electric power. How are the appliances connected between this wire and the ground, in parallel or in series? Explain.

a. The appliances are connected in parallel to provide different voltage differences across each appliance.

b. The appliances are connected in parallel to provide the same voltage difference across each appliance.

c. The appliances are connected in series to provide the same voltage difference across each appliance.

d. The appliances are connected in series to provide different voltage differences across each appliance.

19.4 Electric Power

67. A single resistor is connected across the terminals of a battery When you attach a second resistor in parallel with the first, does the power dissipated by the system change?

a. No, the power dissipated remain same.

b. Yes, the power dissipated increases.

c. Yes, the power dissipated decreases.

68. In a flashlight, the batteries are normally connected in series. Why are they not connected in parallel?

a. Batteries are connected in series for higher voltage and power output.

b. Batteries are connected in series for lower voltage and power output.

c. Batteries are connected in series so that power output is a much lower for the same amount of voltage.

d. Batteries are connected in series to reduce the overall loss of energy from the circuit.

CHAPTER 20
Magnetism

Figure 20.1 The magnificent spectacle of the Aurora Borealis, or northern lights, glows in the northern sky above Bear Lake near Eielson Air Force Base, Alaska. Shaped by Earth's magnetic field, this light is produced by radiation spewed from solar storms. (credit: Senior Airman Joshua Strang, Flickr)

Chapter Outline

20.1 Magnetic Fields, Field Lines, and Force

20.2 Motors, Generators, and Transformers

20.3 Electromagnetic Induction

INTRODUCTION You may have encountered magnets for the first time as a small child playing with magnetic toys or refrigerator magnets. At the time, you likely noticed that two magnets that repulse each other will attract each other if you flip one of them around. The force that acts across the air gaps between magnets is the same force that creates wonders such as the Aurora Borealis. In fact, magnetic effects pervade our lives in myriad ways, from electric motors to medical imaging and computer memory. In this chapter, we introduce magnets and learn how they work and how magnetic fields and electric currents interact.

20.1 Magnetic Fields, Field Lines, and Force

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Summarize properties of magnets and describe how some nonmagnetic materials can become magnetized
- Describe and interpret drawings of magnetic fields around permanent magnets and current-carrying wires
- Calculate the magnitude and direction of magnetic force in a magnetic field and the force on a current-carrying wire in a magnetic field

Section Key Terms

Curie temperature	domain	electromagnet	electromagnetism	ferromagnetic
magnetic dipole	magnetic field	magnetic pole	magnetized	north pole
permanent magnet	right-hand rule	solenoid	south pole	

Magnets and Magnetization

People have been aware of magnets and magnetism for thousands of years. The earliest records date back to ancient times, particularly in the region of Asia Minor called Magnesia—the name of this region is the source of words like *magnet*. Magnetic rocks found in Magnesia, which is now part of western Turkey, stimulated interest during ancient times. When humans first discovered magnetic rocks, they likely found that certain parts of these rocks attracted bits of iron or other magnetic rocks more strongly than other parts. These areas are called the *poles* of a magnet. A **magnetic pole** is the part of a magnet that exerts the strongest force on other magnets or magnetic material, such as iron. For example, the poles of the bar magnet shown in Figure 20.2 are where the paper clips are concentrated.

Figure 20.2 A bar magnet with paper clips attracted to the two poles.

If a bar magnet is suspended so that it rotates freely, one pole of the magnet will always turn toward the north, with the opposite pole facing south. This discovery led to the compass, which is simply a small, elongated magnet mounted so that it can rotate freely. An example of a compass is shown Figure 20.3. The pole of the magnet that orients northward is called the **north pole**, and the opposite pole of the magnet is called the **south pole**.

Figure 20.3 A compass is an elongated magnet mounted in a device that allows the magnet to rotate freely.

The discovery that one particular pole of a magnet orients northward, whereas the other pole orients southward allowed people to identify the north and south poles of any magnet. It was then noticed that the north poles of two different magnets repel each other, and likewise for the south poles. Conversely, the north pole of one magnet attracts the south pole of other magnets. This situation is analogous to that of electric charge, where like charges repel and unlike charges attract. In magnets, we simply replace charge with *pole*: Like poles repel and unlike poles attract. This is summarized in Figure 20.4, which shows how the force between magnets depends on their relative orientation.

Figure 20.4 Depending on their relative orientation, magnet poles will either attract each other or repel each other.

Consider again the fact that the pole of a magnet that orients northward is called the north pole of the magnet. If unlike poles attract, then the magnetic pole of Earth that is close to the geographic North Pole must be a magnetic south pole! Likewise, the magnetic pole of Earth that is close to the geographic South Pole must be a magnetic north pole. This situation is depicted in Figure 20.5, in which Earth is represented as containing a giant internal bar magnet with its magnetic south pole at the geographic North Pole and vice versa. If we were to somehow suspend a giant bar magnet in space near Earth, then the north pole of the space magnet would be attracted to the south pole of Earth's internal magnet. This is in essence what happens with a compass needle: Its magnetic north pole is attracted to the magnet south pole of Earth's internal magnet.

Figure 20.5 Earth can be thought of as containing a giant magnet running through its core. The magnetic south pole of Earth's magnet is at the geographic North Pole, so the north pole of magnets is attracted to the North Pole, which is how the north pole of magnets got their name. Likewise, the south pole of magnets is attracted to the geographic South Pole of Earth.

What happens if you cut a bar magnet in half? Do you obtain one magnet with two south poles and one magnet with two north poles? The answer is no: Each half of the bar magnet has a north pole and a south pole. You can even continue cutting each piece of the bar magnet in half, and you will always obtain a new, smaller magnet with two opposite poles. As shown in Figure 20.6, you can continue this process down to the atomic scale, and you will find that even the smallest particles that behave as magnets have two opposite poles. In fact, no experiment has ever found any object with a single magnetic pole, from the smallest subatomic particle such as electrons to the largest objects in the universe such as stars. Because magnets always have two poles, they are referred to as **magnetic dipoles**—*di* means *two*. Below, we will see that magnetic dipoles have properties that are analogous to electric dipoles.

Figure 20.6 All magnets have two opposite poles, from the smallest, such as subatomic particles, to the largest, such as stars.

WATCH PHYSICS

Introduction to Magnetism

This video provides an interesting introduction to magnetism and discusses, in particular, how electrons around their atoms contribute to the magnetic effects that we observe.

Click to view content (https://www.openstax.org/l/28_intro_magn)

GRASP CHECK

Toward which magnetic pole of Earth is the north pole of a compass needle attracted?

a. The north pole of a compass needle is attracted to the north magnetic pole of Earth, which is located near the geographic North Pole of Earth.
b. The north pole of a compass needle is attracted to the south magnetic pole of Earth, which is located near the geographic North Pole of Earth.
c. The north pole of a compass needle is attracted to the north magnetic pole of Earth, which is located near the geographic South Pole of Earth.
d. The north pole of a compass needle is attracted to the south magnetic pole of Earth, which is located near the geographic South Pole of Earth.

Only certain materials, such as iron, cobalt, nickel, and gadolinium, exhibit strong magnetic effects. Such materials are called **ferromagnetic**, after the Latin word *ferrum* for iron. Other materials exhibit weak magnetic effects, which are detectable only with sensitive instruments. Not only do ferromagnetic materials respond strongly to magnets—the way iron is attracted to magnets—but they can also be **magnetized** themselves—that is, they can be induced to be magnetic or made into permanent magnets (Figure 20.7). A **permanent magnet** is simply a material that retains its magnetic behavior for a long time, even when exposed to demagnetizing influences.

Figure 20.7 An unmagnetized piece of iron is placed between two magnets, heated, and then cooled, or simply tapped when cold. The iron becomes a permanent magnet with the poles aligned as shown: Its south pole is adjacent to the north pole of the original magnet, and its north pole is adjacent to the south pole of the original magnet. Note that attractive forces are created between the central magnet and the outer magnets.

When a magnet is brought near a previously unmagnetized ferromagnetic material, it causes local magnetization of the material with unlike poles closest, as in the right side of Figure 20.7. This causes an attractive force, which is why unmagnetized iron is attracted to a magnet.

What happens on a microscopic scale is illustrated in Figure 7(a). Regions within the material called **domains** act like small bar magnets. Within domains, the magnetic poles of individual atoms are aligned. Each atom acts like a tiny bar magnet. Domains are small and randomly oriented in an unmagnetized ferromagnetic object. In response to an external magnetic field, the domains may grow to millimeter size, aligning themselves, as shown in Figure 7(b). This induced magnetization can be made permanent if the material is heated and then cooled, or simply tapped in the presence of other magnets.

Figure 20.8 (a) An unmagnetized piece of iron—or other ferromagnetic material—has randomly oriented domains. (b) When magnetized by an external magnet, the domains show greater alignment, and some grow at the expense of others. Individual atoms are aligned within

domains; each atom acts like a tiny bar magnet.

Conversely, a permanent magnet can be demagnetized by hard blows or by heating it in the absence of another magnet. Increased thermal motion at higher temperature can disrupt and randomize the orientation and size of the domains. There is a well-defined temperature for ferromagnetic materials, which is called the **Curie temperature**, above which they cannot be magnetized. The Curie temperature for iron is 1,043 K (770 °C), which is well above room temperature. There are several elements and alloys that have Curie temperatures much lower than room temperature and are ferromagnetic only below those temperatures.

Snap Lab

Refrigerator Magnets

We know that like magnetic poles repel and unlike poles attract. See if you can show this for two refrigerator magnets. Will the magnets stick if you turn them over? Why do they stick to the refrigerator door anyway? What can you say about the magnetic properties of the refrigerator door near the magnet? Do refrigerator magnets stick to metal or plastic spoons? Do they stick to all types of metal?

GRASP CHECK

You have one magnet with the north and south poles labeled. How can you use this magnet to identify the north and south poles of other magnets?

a. If the north pole of a known magnet is repelled by a pole of an unknown magnet on bringing them closer, that pole of unknown magnet is its north pole; otherwise, it is its south pole.
b. If the north pole of known magnet is attracted to a pole of an unknown magnet on bringing them closer, that pole of unknown magnet is its north pole; otherwise, it is its south pole.

Magnetic Fields

We have thus seen that forces can be applied between magnets and between magnets and ferromagnetic materials without any contact between the objects. This is reminiscent of electric forces, which also act over distances. Electric forces are described using the concept of the electric field, which is a force field around electric charges that describes the force on any other charge placed in the field. Likewise, a magnet creates a **magnetic field** around it that describes the force exerted on other magnets placed in the field. As with electric fields, the pictorial representation of magnetic field lines is very useful for visualizing the strength and direction of the magnetic field.

As shown in Figure 20.9, the direction of magnetic field lines is defined to be the direction in which the north pole of a compass needle points. If you place a compass near the north pole of a magnet, the north pole of the compass needle will be repelled and point away from the magnet. Thus, the magnetic field lines point away from the north pole of a magnet and toward its south pole.

Figure 20.9 The black lines represent the magnetic field lines of a bar magnet. The field lines point in the direction that the north pole of a small compass would point, as shown at left. Magnetic field lines never stop, so the field lines actually penetrate the magnet to form complete loops, as shown at right.

Magnetic field lines can be mapped out using a small compass. The compass is moved from point to point around a magnet, and at each point, a short line is drawn in the direction of the needle, as shown in Figure 20.10. Joining the lines together then reveals the path of the magnetic field line. Another way to visualize magnetic field lines is to sprinkle iron filings around a magnet. The filings will orient themselves along the magnetic field lines, forming a pattern such as that shown on the right in Figure 20.10.

Virtual Physics

Using a Compass to Map Out the Magnetic Field

Click to view content (http://www.openstax.org/l/28magcomp)

This simulation presents you with a bar magnet and a small compass. Begin by dragging the compass around the bar magnet to see in which direction the magnetic field points. Note that the strength of the magnetic field is represented by the brightness of the magnetic field icons in the grid pattern around the magnet. Use the magnetic field meter to check the field strength at several points around the bar magnet. You can also flip the polarity of the magnet, or place Earth on the image to see how the compass orients itself.

GRASP CHECK

With the slider at the top right of the simulation window, set the magnetic field strength to 100 percent . Now use the magnetic field meter to answer the following question: Near the magnet, where is the magnetic field strongest and where is it weakest? Don't forget to check inside the bar magnet.

a. The magnetic field is strongest at the center and weakest between the two poles just outside the bar magnet. The magnetic field lines are densest at the center and least dense between the two poles just outside the bar magnet.
b. The magnetic field is strongest at the center and weakest between the two poles just outside the bar magnet. The magnetic field lines are least dense at the center and densest between the two poles just outside the bar magnet.
c. The magnetic field is weakest at the center and strongest between the two poles just outside the bar magnet. The magnetic field lines are densest at the center and least dense between the two poles just outside the bar magnet.
d. The magnetic field is weakest at the center and strongest between the two poles just outside the bar magnet and the magnetic field lines are least dense at the center and densest between the two poles just outside the bar magnet.

Figure 20.10 Magnetic field lines can be drawn by moving a small compass from point to point around a magnet. At each point, draw a short line in the direction of the compass needle. Joining the points together reveals the path of the magnetic field lines. Another way to visualize magnetic field lines is to sprinkle iron filings around a magnet, as shown at right.

When two magnets are brought close together, the magnetic field lines are perturbed, just as happens for electric field lines when two electric charges are brought together. Bringing two north poles together—or two south poles—will cause a repulsion, and the magnetic field lines will bend away from each other. This is shown in Figure 20.11, which shows the magnetic field lines created by the two closely separated north poles of a bar magnet. When opposite poles of two magnets are brought together, the

magnetic field lines join together and become denser between the poles. This situation is shown in Figure 20.11.

Figure 20.11 (a) When two north poles are approached together, the magnetic field lines repel each other and the two magnets experience a repulsive force. The same occurs if two south poles are approached together. (b) If opposite poles are approached together, the magnetic field lines become denser between the poles and the magnets experience an attractive force.

Like the electric field, the magnetic field is stronger where the lines are denser. Thus, between the two north poles in Figure 20.11, the magnetic field is very weak because the density of the magnetic field is almost zero. A compass placed at that point would essentially spin freely if we ignore Earth's magnetic field. Conversely, the magnetic field lines between the north and south poles in Figure 20.11 are very dense, indicating that the magnetic field is very strong in this region. A compass placed here would quickly align with the magnetic field and point toward the south pole on the right.

Note that magnets are not the only things that make magnetic fields. Early in the nineteenth century, people discovered that electrical currents cause magnetic effects. The first significant observation was by the Danish scientist Hans Christian Oersted (1777–1851), who found that a compass needle was deflected by a current-carrying wire. This was the first significant evidence that the movement of electric charges had any connection with magnets. An **electromagnet** is a device that uses electric current to make a magnetic field. These temporarily induced magnets are called electromagnets. Electromagnets are employed for everything from a wrecking yard crane that lifts scrapped cars to controlling the beam of a 90-km-circumference particle accelerator to the magnets in medical-imaging machines (see Figure 20.12).

Figure 20.12 Instrument for magnetic resonance imaging (MRI). The device uses a cylindrical-coil electromagnet to produce for the main magnetic field. The patient goes into the *tunnel* on the gurney. (credit: Bill McChesney, Flickr)

The magnetic field created by an electric current in a long straight wire is shown in Figure 20.13. The magnetic field lines form concentric circles around the wire. The direction of the magnetic field can be determined using the *right-hand rule*. This rule shows up in several places in the study of electricity and magnetism. Applied to a straight current-carrying wire, the **right-hand rule** says that, with your right thumb pointed in the direction of the current, the magnetic field will be in the direction in which your right fingers curl, as shown in Figure 20.13. If the wire is very long compared to the distance r from the wire, the strength B of the magnetic field is given by

$$B_{\text{straightwire}} = \frac{\mu_0 I}{2\pi r} \tag{20.1}$$

where I is the current in the wire in amperes. The SI unit for magnetic field is the tesla (T). The symbol μ_0 —read "mu-zero"—is a constant called the "permeability of free space" and is given by

$$\mu_0 = 4\pi \times 10^{-7}\,\text{T}\cdot\text{m/A}. \tag{20.2}$$

Figure 20.13 This image shows how to use the right-hand rule to determine the direction of the magnetic field created by current flowing through a straight wire. Point your right thumb in the direction of the current, and the magnetic field will be in the direction in which your fingers curl.

WATCH PHYSICS

Magnetic Field Due to an Electric Current

This video describes the magnetic field created by a straight current-carrying wire. It goes over the right-hand rule to determine the direction of the magnetic field, and presents and discusses the formula for the strength of the magnetic field due to a straight current-carrying wire.

Click to view content (https://www.openstax.org/l/28magfield)

GRASP CHECK

A long straight wire is placed on a table top and electric current flows through the wire from right to left. If you look at the wire end-on from the left end, does the magnetic field go clockwise or counterclockwise?

a. By pointing your right-hand thumb in the direction opposite of current, the right-hand fingers will curl counterclockwise, so the magnetic field will be in the counterclockwise direction.
b. By pointing your right-hand thumb in the direction opposite of current, the right-hand fingers will curl clockwise, so the magnetic field will be in the clockwise direction.
c. By pointing your right-hand thumb in the direction of current, the right-hand fingers will curl counterclockwise, so the magnetic field will be in the counterclockwise direction.
d. By pointing your right-hand thumb in the direction of current, the right-hand fingers will curl clockwise, so the magnetic field will be in the clockwise direction.

Now imagine winding a wire around a cylinder with the cylinder then removed. The result is a wire coil, as shown in Figure 20.14. This is called a **solenoid**. To find the direction of the magnetic field produced by a solenoid, apply the right-hand rule to several points on the coil. You should be able to convince yourself that, inside the coil, the magnetic field points from left to right. In fact, another application of the right-hand rule is to curl your right-hand fingers around the coil in the direction in which the current flows. Your right thumb then points in the direction of the magnetic field inside the coil: left to right in this case.

Figure 20.14 A wire coil with current running through as shown produces a magnetic field in the direction of the red arrow.

Each loop of wire contributes to the magnetic field inside the solenoid. Because the magnetic field lines must form closed loops, the field lines close the loop outside the solenoid. The magnetic field lines are much denser inside the solenoid than outside the solenoid. The resulting magnetic field looks very much like that of a bar magnet, as shown in Figure 20.15. The magnetic field strength deep inside a solenoid is

$$B_{\text{solenoid}} = \mu_0 \frac{NI}{\ell}, \tag{20.3}$$

where N is the number of wire loops in the solenoid and ℓ is the length of the solenoid.

Figure 20.15 Iron filings show the magnetic field pattern around (a) a solenoid and (b) a bar magnet. The fields patterns are very similar, especially near the ends of the solenoid and bar magnet.

Virtual Physics

Electromagnets

Click to view content (http://www.openstax.org/l/28elec_magnet)

Use this simulation to visualize the magnetic field made from a solenoid. Be sure to click on the tab that says Electromagnet. You can drive AC or DC current through the solenoid by choosing the appropriate current source. Use the field meter to measure the strength of the magnetic field and then change the number of loops in the solenoid to see how this affects the magnetic field strength.

GRASP CHECK

Choose the battery as current source and set the number of wire loops to four. With a nonzero current going through the solenoid, measure the magnetic field strength at a point. Now decrease the number of wire loops to two. How does the magnetic field strength change at the point you chose?

a. There will be no change in magnetic field strength when number of loops reduces from four to two.
b. The magnetic field strength decreases to half of its initial value when number of loops reduces from four to two.
c. The magnetic field strength increases to twice of its initial value when number of loops reduces from four to two.
d. The magnetic field strength increases to four times of its initial value when number of loops reduces from four to two.

Magnetic Force

If a moving electric charge, that is electric current, produces a magnetic field that can exert a force on another magnet, then the reverse should be true by Newton's third law. In other words, a charge moving through the magnetic field produced by another object should experience a force—and this is exactly what we find. As a concrete example, consider Figure 20.16, which shows a

charge q moving with velocity $\vec{v}$ through a magnetic field $\vec{B}$ between the poles of a permanent magnet. The magnitude F of the force experienced by this charge is

$$F = qvB\sin\theta, \tag{20.4}$$

where θ is the angle between the velocity of the charge and the magnetic field.

The direction of the force may be found by using another version of the right-hand rule: First, we join the tails of the velocity vector and a magnetic field vector, as shown in step 1 of Figure 20.16. We then curl our right fingers from $\vec{v}$ to $\vec{B}$, as indicated in step (2) of Figure 20.16. The direction in which the right thumb points is the direction of the force. For the charge in Figure 20.16, we find that the force is directed into the page.

Note that the factor $\sin\theta$ in the equation $F = qvB\sin\theta$ means that zero force is applied on a charge that moves parallel to a magnetic field because $\theta = 0$ and $\sin 0 = 0$. The maximum force a charge can experience is when it moves perpendicular to the magnetic field, because $\theta = 90°$ and $\sin 90° = 1$.

Figure 20.16 (a) An electron moves through a uniform magnetic field. (b) Using the right-hand rule, the force on the electron is found to be directed into the page.

LINKS TO PHYSICS

Magnetohydrodynamic Drive

In Tom Clancy's Cold War novel "The Hunt for Red October," the Soviet Union built a submarine (see Figure 20.17) with a magnetohydrodynamic drive that was so silent it could not be detected by surface ships. The only conceivable purpose to build such a submarine was to give the Soviet Union first-strike capability, because this submarine could sneak close to the coast of the United States and fire its ballistic missiles, destroying key military and government installations to prevent an American counterattack.

Figure 20.17 A Typhoon-class Russian ballistic-missile submarine on which the fictional submarine Red October was based.

A magnetohydrodynamic drive is supposed to be silent because it has no moving parts. Instead, it uses the force experienced by charged particles that move in a magnetic field. The basic idea behind such a drive is depicted in Figure 20.18. Salt water flows through a channel that runs from the front to the back of the submarine. A magnetic field is applied horizontally across the channel, and a voltage is applied across the electrodes on the top and bottom of the channel to force a downward electric current through the water. The charge carriers are the positive sodium ions and the negative chlorine ions of salt. Using the right-hand

rule, the force on the charge carriers is found to be toward the rear of the vessel. The accelerated charges collide with water molecules and transfer their momentum, creating a jet of water that is propelled out the rear of the channel. By Newton's third law, the vessel experiences a force of equal magnitude, but in the opposite direction.

Figure 20.18 A schematic drawing of a magnetohydrodynamic drive showing the water channel, the current direction, the magnetic field direction, and the resulting force.

Fortunately for all involved, it turns out that such a propulsion system is not very practical. Some back-of-the-envelope calculations show that, to power a submarine, either extraordinarily high magnetic fields or extraordinarily high electric currents would be required to obtain a reasonable thrust. In addition, prototypes of magnetohydrodynamic drives show that they are anything but silent. Electrolysis caused by running a current through salt water creates bubbles of hydrogen and oxygen, which makes this propulsion system quite noisy. The system also leaves a trail of chloride ions and metal chlorides that can easily be detected to locate the submarine. Finally, the chloride ions are extremely reactive and very quickly corrode metal parts, such as the electrode or the water channel itself. Thus, the Red October remains in the realm of fiction, but the physics involved is quite real.

GRASP CHECK

If the magnetic field is downward, in what direction must the current flow to obtain rearward-pointing force?

a. The current must flow vertically from up to down when viewed from the rear of the boat.
b. The current must flow vertically from down to up when viewed from the rear of the boat.
c. The current must flow horizontally from left to right when viewed from the rear of the boat.
d. The current must flow horizontally from right to left when viewed from the rear of the boat.

Instead of a single charge moving through a magnetic field, consider now a steady current I moving through a straight wire. If we place this wire in a uniform magnetic field, as shown in Figure 20.19, what is the force on the wire or, more precisely, on the electrons in the wire? An electric current involves charges that move. If the charges q move a distance ℓ in a time t, then their speed is $v = \ell/t$. Inserting this into the equation $F = qvB\sin\theta$ gives

$$\begin{aligned} F &= q\left(\frac{\ell}{t}\right)B\sin\theta \\ &= \left(\frac{q}{t}\right)\ell B\sin\theta. \end{aligned} \tag{20.5}$$

The factor q/t in this equation is nothing more than the current in the wire. Thus, using $I = q/t$, we obtain

$$F = I\ell B\sin\theta(1.4). \tag{20.6}$$

This equation gives the force on a straight current-carrying wire of length ℓ in a magnetic field of strength B. The angle θ is the angle between the current vector and the magnetic field vector. Note that ℓ is the length of wire that is in the magnetic field and for which $\theta \neq 0$, as shown in Figure 20.19.

The direction of the force is determined in the same way as for a single charge. Curl your right fingers from the vector for I to the vector for B, and your right thumb will point in the direction of the force on the wire. For the wire shown in Figure 20.19, the force is directed into the page.

Figure 20.19 A straight wire carrying current I in a magnetic field B. The force exerted on the wire is directed into the page. The length ℓ is the length of the wire that is *in* the magnetic field.

Throughout this section, you may have noticed the symmetries between magnetic effects and electric effects. These effects all fall under the umbrella of **electromagnetism**, which is the study of electric and magnetic phenomena. We have seen that electric charges produce electric fields, and moving electric charges produce magnetic fields. A magnetic dipole produces a magnetic field, and, as we will see in the next section, moving magnetic dipoles produce an electric field. Thus, electricity and magnetism are two intimately related and symmetric phenomena.

WORKED EXAMPLE

Trajectory of Electron in Magnetic Field

A proton enters a region of constant magnetic field, as shown in Figure 20.20. The magnetic field is coming out of the page. If the electron is moving at 3.0×10^6 m/s and the magnetic field strength is 2.0 T, what is the magnitude and direction of the force on the proton?

Figure 20.20 A proton enters a region of uniform magnetic field. The magnetic field is coming out of the page—the circles with dots represent vector arrow heads coming out of the page.

STRATEGY

Use the equation $F = qvB\sin\theta$ to find the magnitude of the force on the proton. The angle between the magnetic field vectors and the velocity vector of the proton is $90°$. The direction of the force may be found by using the right-hand rule.

Solution

The charge of the proton is $q = 1.60 \times 10^{-19}\text{C}$. Entering this value and the given velocity and magnetic field strength into the equation $F = qvB\sin\theta$ gives

$$\begin{aligned} F &= qvB\sin\theta \\ &= \left(1.60 \times 10^{-19}\text{C}\right)\left(3.0 \times 10^6\text{m/s}\right)(2.0\text{ T})\sin(90°) \\ &= 9.6 \times 10^{-13}\text{N}. \end{aligned} \tag{20.7}$$

To find the direction of the force, first join the velocity vector end to end with the magnetic field vector, as shown in Figure 20.21. Now place your right hand so that your fingers point in the direction of the velocity and curl them upward toward the magnetic field vector. The force is in the direction in which your thumb points. In this case, the force is downward in the plane of the paper in the $\hat{z}$ -direction, as shown in Figure 20.21.

Figure 20.21 The velocity vector and a magnetic field vector from Figure 20.20 are placed end to end. A right hand is shown with the fingers curling up from the velocity vector toward the magnetic field vector. The thumb points in the direction of the resulting force, which is the $\hat{z}$ -direction in this case.

Thus, combining the magnitude and the direction, we find that the force on the proton is $\left(9.6 \times 10^{-13}\text{N}\right)\hat{z}$.

Discussion

This seems like a very small force. However, the proton has a mass of $1.67 \times 10^{-27}\text{kg}$, so its acceleration is $a = \frac{F}{m} = \frac{9.6 \times 10^{-13}\text{N}}{1.67 \times 10^{-27}\text{kg}} = 5.7 \times 10^{14}\ \text{m/s}^2$, or about ten thousand billion times the acceleration due to gravity!

We found that the proton's initial acceleration as it enters the magnetic field is downward in the plane of the page. Notice that, as the proton accelerates, its velocity remains perpendicular to the magnetic field, so the magnitude of the force does not change. In addition, because of the right-hand rule, the direction of the force remains perpendicular to the velocity. This force is nothing more than a centripetal force: It has a constant magnitude and is always perpendicular to the velocity. Thus, the magnitude of the velocity does not change, and the proton executes circular motion. The radius of this circle may be found by using the kinematics relationship.

$$\begin{aligned} F &= ma = m\frac{v^2}{r} \\ a &= \frac{v^2}{r} \\ r &= \frac{v^2}{a} = \frac{\left(3.0 \times 10^6\text{m/s}\right)^2}{5.7 \times 10^{14}\ \text{m/s}^2} = 1.6\ \text{cm} \end{aligned} \tag{20.8}$$

The path of the proton in the magnetic field is shown in Figure 20.22.

Figure 20.22 When traveling perpendicular to a constant magnetic field, a charged particle will execute circular motion, as shown here for a proton.

WORKED EXAMPLE

Wire with Current in Magnetic Field

Now suppose we run a wire through the uniform magnetic field from the previous example, as shown. If the wire carries a current of 1.0 A in the $\hat{y}$ -direction, and the region with magnetic field is 4.0 cm long, what is the force on the wire?

STRATEGY

Use equation $F = I\ell B \sin\theta$ to find the magnitude of the force on the wire. The length of the wire inside the magnetic field is 4.0 cm, and the angle between the current direction and the magnetic field direction is 90°. To find the direction of the force, use the right-hand rule as explained just after the equation $F = I\ell B \sin\theta$.

Solution

Insert the given values into equation $F = I\ell B \sin\theta$ to find the magnitude of the force

$$F = I\ell B \sin\theta = (1.5\ \text{A})\,(0.040\ \text{m})\,(2.0\ \text{T}) = 0.12\ \text{N}. \tag{20.9}$$

To find the direction of the force, begin by placing the current vector end to end with a vector for the magnetic field. The result is as shown in the figure in the previous Worked Example with $\vec{v}$ replaced by $\vec{I}$. Curl your right-hand fingers from $\vec{I}$ to $\vec{B}$ and your right thumb points down the page, again as shown in the figure in the previous Worked Example. Thus, the direction of the force is in the $\hat{x}$-direction. The complete force is thus $(0.12\ \text{N})\,\hat{x}$.

Discussion

The direction of the force is the same as the initial direction of the force was in the previous example for a proton. However, because the current in a wire is confined to a wire, the direction in which the charges move does not change. Instead, the entire wire accelerates in the $\hat{x}$-direction. The force on a current-carrying wire in a magnetic field is the basis of all electrical motors, as we will see in the upcoming sections.

Practice Problems

1. What is the magnitude of the force on an electron moving at 1.0 × 106 m/s perpendicular to a 1.0-T magnetic field?
   a. 0.8×10^{-13} N
   b. 1.6×10^{-14} N
   c. 0.8×10^{-14} N
   d. 1.6×10^{-13} N

2. A straight 10 cm wire carries 0.40 A and is oriented perpendicular to a magnetic field. If the force on the wire is 0.022 N, what is the magnitude of the magnetic field?
   a. 1.10×10^{-2} T
   b. 0.55×10^{-2} T
   c. 1.10 T
   d. 0.55 T

Check Your Understanding

3. If two magnets repel each other, what can you conclude about their relative orientation?
   a. Either the south pole of magnet 1 is closer to the north pole of magnet 2 or the north pole of magnet 1 is closer to the south pole of magnet 2.
   b. Either the south poles of both the magnet 1 and magnet 2 are closer to each other or the north poles of both the magnet 1

and magnet 2 are closer to each other.

4. Describe methods to demagnetize a ferromagnet.
   a. by cooling, heating, or submerging in water
   b. by heating, hammering, and spinning it in external magnetic field
   c. by hammering, heating, and rubbing with cloth
   d. by cooling, submerging in water, or rubbing with cloth

5. What is a magnetic field?
   a. The directional lines present inside and outside the magnetic material that indicate the magnitude and direction of the magnetic force.
   b. The directional lines present inside and outside the magnetic material that indicate the magnitude of the magnetic force.
   c. The directional lines present inside the magnetic material that indicate the magnitude and the direction of the magnetic force.
   d. The directional lines present outside the magnetic material that indicate the magnitude and the direction of the magnetic force.

6. Which of the following drawings is correct?

a.

b.

c.

d.

20.2 Motors, Generators, and Transformers

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Explain how electric motors, generators, and transformers work
- Explain how commercial electric power is produced, transmitted, and distributed

Section Key Terms

electric motor	generator	transformer

Electric Motors, Generators, and Transformers

As we learned previously, a current-carrying wire in a magnetic field experiences a force—recall $F = I\ell B \sin\theta$. **Electric motors**, which convert electrical energy into mechanical energy, are the most common application of magnetic force on current-carrying wires. Motors consist of loops of wire in a magnetic field. When current is passed through the loops, the magnetic field exerts a torque on the loops, which rotates a shaft. Electrical energy is converted to mechanical work in the process. Figure 20.23 shows a schematic drawing of an electric motor.

Figure 20.23 Torque on a current loop. A vertical loop of wire in a horizontal magnetic field is attached to a vertical shaft. When current is passed through the wire loop, torque is exerted on it, making it turn the shaft.

Let us examine the force on each segment of the loop in Figure 20.23 to find the torques produced about the axis of the vertical shaft—this will lead to a useful equation for the torque on the loop. We take the magnetic field to be uniform over the rectangular loop, which has width w and height ℓ, as shown in the figure. First, consider the force on the top segment of the loop. To determine the direction of the force, we use the right-hand rule. The current goes from left to right into the page, and the magnetic field goes from left to right in the plane of the page. Curl your right fingers from the current vector to the magnetic field vector and your right thumb points down. Thus, the force on the top segment is downward, which produces no torque on the shaft. Repeating this analysis for the bottom segment—neglect the small gap where the lead wires go out—shows that the force on the bottom segment is upward, again producing no torque on the shaft.

Consider now the left vertical segment of the loop. Again using the right-hand rule, we find that the force exerted on this segment is perpendicular to the magnetic field, as shown in Figure 20.23. This force produces a torque on the shaft. Repeating this analysis on the right vertical segment of the loop shows that the force on this segment is in the direction opposite that of the force on the left segment, thereby producing an equal torque on the shaft. The total torque on the shaft is thus twice the toque on one of the vertical segments of the loop.

To find the magnitude of the torque as the wire loop spins, consider Figure 20.24, which shows a view of the wire loop from above. Recall that torque is defined as $\tau = rF\sin\theta$, where F is the applied force, r is the distance from the pivot to where the force is applied, and θ is the angle between r and F. Notice that, as the loop spins, the current in the vertical loop segments is always perpendicular to the magnetic field. Thus, the equation $F = I\ell B\sin\theta$ gives the magnitude of the force on each vertical segment as $F = I\ell B$. The distance r from the shaft to where this force is applied is $w/2$, so the torque created by this force is

$$\tau_{\text{segment}} = rF\sin\theta = w/2 I\ell B\sin\theta = (w/2)\, I\ell B\sin\theta. \tag{20.10}$$

Because there are two vertical segments, the total torque is twice this, or

$$\tau = wI\ell B\sin\theta. \tag{20.11}$$

If we have a multiple loop with N turns, we get N times the torque of a single loop. Using the fact that the area of the loop is $A = w\ell$; the expression for the torque becomes

$$\tau = NIAB \sin\theta. \tag{20.12}$$

This is the torque on a current-carrying loop in a uniform magnetic field. This equation can be shown to be valid for a loop of any shape.

Figure 20.24 View from above of the wire loop from Figure 20.23. The magnetic field generates a force *F* on each vertical segment of the wire loop, which generates a torque on the shaft. Notice that the currents I_{in} and I_{out} have the same magnitude because they both represent the current flowing in the wire loop, but I_{in} flows into the page and I_{out} flows out of the page.

From the equation $\tau = NIAB \sin\theta$, we see that the torque is zero when $\theta = 0$. As the wire loop rotates, the torque increases to a maximum positive torque of $w\ell B$ when $\theta = 90°$. The torque then decreases back to zero as the wire loop rotates to $\theta = 180°$. From $\theta = 180°$ to $\theta = 360°$, the torque is negative. Thus, the torque changes sign every half turn, so the wire loop will oscillate back and forth.

For the coil to continue rotating in the same direction, the current is reversed as the coil passes through $\theta = 0$ and $\theta = 180°$ using automatic switches called *brushes*, as shown in Figure 20.25.

Figure 20.25 (a) As the angular momentum of the coil carries it through $\theta = 0$, the brushes reverse the current and the torque remains clockwise. (b) The coil rotates continuously in the clockwise direction, with the current reversing each half revolution to maintain the clockwise torque.

Consider now what happens if we run the motor in reverse; that is, we attach a handle to the shaft and mechanically force the coil to rotate within the magnetic field, as shown in Figure 20.26. As per the equation $F = qvB \sin\theta$—where θ is the angle between the vectors $\vec{v}$ and $\vec{B}$—charges in the wires of the loop experience a magnetic force because they are moving in a magnetic field. Again using the right-hand rule, where we curl our fingers from vector $\vec{v}$ to vector $\vec{B}$, we find that charges in the top and bottom segments feel a force perpendicular to the wire, which does not cause a current. However, charges in the vertical wires experience forces parallel to the wire, causing a current to flow through the wire and through an external circuit if one is connected. A device such as this that converts mechanical energy into electrical energy is called a **generator**.

Figure 20.26 When this coil is rotated through one-fourth of a revolution, the magnetic flux Φ changes from its maximum to zero, inducing an emf, which drives a current through an external circuit.

Because current is induced only in the side wires, we can find the induced emf by only considering these wires. As explained in Induced Current in a Wire, motional emf in a straight wire moving at velocity v through a magnetic field B is $E = B\ell v$, where the velocity is perpendicular to the magnetic field. In the generator, the velocity makes an angle θ with B (see Figure 20.27), so the velocity component perpendicular to B is $v \sin\theta$. Thus, in this case, the emf induced on each vertical wire segment is $E = B\ell v \sin\theta$, and they are in the same direction. The total emf around the loop is then

$$E = 2B\ell v \sin\theta. \tag{20.13}$$

Although this expression is valid, it does not give the emf as a function of time. To find how the emf evolves in time, we assume that the coil is rotated at a constant angular velocity ω. The angle θ is related to the angular velocity by $\theta = \omega t$, so that

$$E = 2B\ell v \sin\omega t. \tag{20.14}$$

Recall that tangential velocity v is related to angular velocity ω by $v = r\omega$. Here, $r = w/2$, so that $v = (w/2)\,\omega$ and

$$E = 2B\ell\left(\frac{w}{2}\omega\right)\sin\omega t = B\ell w\omega \sin\omega t. \tag{20.15}$$

Noting that the area of the loop is $A = \ell w$ and allowing for N wire loops, we find that

$$E = NAB\omega \sin\omega t \tag{20.16}$$

is the emf induced in a generator coil of N turns and area A rotating at a constant angular velocity ω in a uniform magnetic field B. This can also be expressed as

$$E = E_0 \sin\omega t \tag{20.17}$$

where

$$E_0 = NAB\omega \tag{20.18}$$

is the maximum (peak) emf.

Figure 20.27 The instantaneous velocity of the vertical wire segments makes an angle θ with the magnetic field. The velocity is shown in the figure by the green arrow, and the angle θ is indicated.

Figure 20.28 shows a generator connected to a light bulb and a graph of the emf vs. time. Note that the emf oscillates from a positive maximum of E_0 to a negative maximum of $-E_0$. In between, the emf goes through zero, which means that zero current flows through the light bulb at these times. Thus, the light bulb actually flickers on and off at a frequency of $2f$, because there are two zero crossings per period. Since alternating current such as this is used in homes around the world, why do we not notice the lights flickering on and off? In the United States, the frequency of alternating current is 60 Hz, so the lights flicker on and off at a frequency of 120 Hz. This is faster than the refresh rate of the human eye, so you don't notice the flicker of the lights. Also, other factors prevent various different types of light bulbs from switching on and off so fast, so the light output is *smoothed out* a bit.

Figure 20.28 The emf of a generator is sent to a light bulb with the system of rings and brushes shown. The graph gives the emf of the generator as a function of time. E_0 is the peak emf. The period is $T = 1/f = 2\pi/\omega$, where f is the frequency at which the coil is rotated in the magnetic field.

Virtual Physics

Generator

Click to view content (http://www.openstax.org/l/28gen)

Use this simulation to discover how an electrical generator works. Control the water supply that makes a water wheel turn a magnet. This induces an emf in a nearby wire coil, which is used to light a light bulb. You can also replace the light bulb with a voltmeter, which allows you to see the polarity of the voltage, which changes from positive to negative.

GRASP CHECK

Set the number of wire loops to three, the bar-magnet strength to about 50 percent, and the loop area to 100 percent. Note the maximum voltage on the voltmeter. Assuming that one major division on the voltmeter is 5V, what is the maximum voltage when using only a single wire loop instead of three wire loops?

a. 5 V
b. 15 V

c. 125 V
d. 53 V

In real life, electric generators look a lot different than the figures in this section, but the principles are the same. The source of mechanical energy that turns the coil can be falling water—hydropower—steam produced by the burning of fossil fuels, or the kinetic energy of wind. Figure 20.29 shows a cutaway view of a steam turbine; steam moves over the blades connected to the shaft, which rotates the coil within the generator.

Figure 20.29 Steam turbine generator. The steam produced by burning coal impacts the turbine blades, turning the shaft which is connected to the generator. (credit: Nabonaco, Wikimedia Commons)

Another very useful and common device that exploits magnetic induction is called a **transformer**. Transformers do what their name implies—they transform voltages from one value to another; the term voltage is used rather than emf because transformers have internal resistance. For example, many cell phones, laptops, video games, power tools, and small appliances have a transformer built into their plug-in unit that changes 120 V or 240 V AC into whatever voltage the device uses. Figure 20.30 shows two different transformers. Notice the wire coils that are visible in each device. The purpose of these coils is explained below.

Figure 20.30 On the left is a common laminated-core transformer, which is widely used in electric power transmission and electrical appliances. On the right is a toroidal transformer, which is smaller than the laminated-core transformer for the same power rating but is more expensive to make because of the equipment required to wind the wires in the doughnut shape.

Figure 20.31 shows a laminated-coil transformer, which is based on Faraday's law of induction and is very similar in construction to the apparatus Faraday used to demonstrate that magnetic fields can generate electric currents. The two wire coils are called the primary and secondary coils. In normal use, the input voltage is applied across the primary coil, and the secondary produces the transformed output voltage. Not only does the iron core trap the magnetic field created by the primary coil, but also its magnetization increases the field strength, which is analogous to how a dielectric increases the electric field strength in a capacitor. Since the input voltage is AC, a time-varying magnetic flux is sent through the secondary coil, inducing an AC output voltage.

Figure 20.31 A typical construction of a simple transformer has two coils wound on a ferromagnetic core. The magnetic field created by the primary coil is mostly confined to and increased by the core, which transmits it to the secondary coil. Any change in current in the primary coil induces a current in the secondary coil.

LINKS TO PHYSICS

Magnetic Rope Memory

To send men to the moon, the Apollo program had to design an onboard computer system that would be robust, consume little power, and be small enough to fit onboard the spacecraft. In the 1960s, when the Apollo program was launched, entire buildings were regularly dedicated to housing computers whose computing power would be easily outstripped by today's most basic handheld calculator.

To address this problem, engineers at MIT and a major defense contractor turned to *magnetic rope memory*, which was an offshoot of a similar technology used prior to that time for creating random access memories. Unlike random access memory, magnetic rope memory was read-only memory that contained not only data but instructions as well. Thus, it was actually more than memory: It was a hard-wired computer program.

The components of magnetic rope memory were wires and iron rings—which were called *cores*. The iron cores served as transformers, such as that shown in the previous figure. However, instead of looping the wires multiple times around the core, individual wires passed only a single time through the cores, making these single-turn transformers. Up to 63 *word* wires could pass through a single core, along with a single *bit* wire. If a word wire passed through a given core, a voltage pulse on this wire would induce an emf in the bit wire, which would be interpreted as a *one*. If the word wire did not pass through the core, no emf would be induced on the bit wire, which would be interpreted as a *zero*.

Engineers would create programs that would be hard wired into these magnetic rope memories. The wiring process could take as long as a month to complete as workers painstakingly threaded wires through some cores and around others. If any mistakes were made either in the programming or the wiring, debugging would be extraordinarily difficult, if not impossible.

These modules did their job quite well. They are credited with correcting an astronaut mistake in the lunar landing procedure, thereby allowing Apollo 11 to land on the moon. It is doubtful that Michael Faraday ever imagined such an application for magnetic induction when he discovered it.

GRASP CHECK

If the bit wire were looped twice around each core, how would the voltage induced in the bit wire be affected?

a. If number of loops around the wire is doubled, the emf is halved.
b. If number of loops around the wire is doubled, the emf is not affected.
c. If number of loops around the wire is doubled, the emf is also doubled.
d. If number of loops around the wire is doubled, the emf is four times the initial value.

For the transformer shown in Figure 20.31, the output voltage V_S from the secondary coil depends almost entirely on the input voltage V_P across the primary coil and the number of loops in the primary and secondary coils. Faraday's law of induction for the secondary coil gives its induced output voltage V_S to be

$$V_S = -N_S \frac{\Delta \Phi}{\Delta t}, \tag{20.19}$$

where N_S is the number of loops in the secondary coil and $\Delta\Phi/\Delta t$ is the rate of change of magnetic flux. The output voltage equals the induced emf ($V_S = E_S$), provided coil resistance is small—a reasonable assumption for transformers. The cross-sectional area of the coils is the same on each side, as is the magnetic field strength, and so $\Delta\Phi/\Delta t$ is the same on each side. The input primary voltage V_P is also related to changing flux by

$$V_P = -N_P \frac{\Delta\Phi}{\Delta t}. \tag{20.20}$$

Taking the ratio of these last two equations yields the useful relationship

$$\frac{V_S}{V_P} = \frac{N_S}{N_P}(3.07). \tag{20.21}$$

This is known as the transformer equation. It simply states that the ratio of the secondary voltage to the primary voltage in a transformer equals the ratio of the number of loops in secondary coil to the number of loops in the primary coil.

Transmission of Electrical Power

Transformers are widely used in the electric power industry to increase voltages—called *step-up* transformers—before long-distance transmission via high-voltage wires. They are also used to decrease voltages—called *step-down* transformers—to deliver power to homes and businesses. The overwhelming majority of electric power is generated by using magnetic induction, whereby a wire coil or copper disk is rotated in a magnetic field. The primary energy required to rotate the coils or disk can be provided by a variety of means. Hydroelectric power plants use the kinetic energy of water to drive electric generators. Coal or nuclear power plants create steam to drive steam turbines that turn the coils. Other sources of primary energy include wind, tides, or waves on water.

Once power is generated, it must be transmitted to the consumer, which often means transmitting power over hundreds of kilometers. To do this, the voltage of the power plant is increased by a step-up transformer, that is stepped up, and the current decreases proportionally because

$$P_{\text{transmitted}} = I_{\text{transmitted}} V_{\text{transmitted}}. \tag{20.22}$$

The lower current $I_{\text{transmitted}}$ in the transmission wires reduces the *Joule losses*, which is heating of the wire due to a current flow. This heating is caused by the small, but nonzero, resistance R_{wire} of the transmission wires. The power lost to the environment through this heat is

$$P_{\text{lost}} = I^2_{\text{transmitted}} R_{\text{wire}}, \tag{20.23}$$

which is proportional to the current *squared* in the transmission wire. This is why the transmitted current $I_{\text{transmitted}}$ must be as small as possible and, consequently, the voltage must be large to transmit the power $P_{\text{transmitted}}$.

Voltages ranging from 120 to 700 kV are used for transmitting power over long distances. The voltage is stepped up at the exit of the power station by a step-up transformer, as shown in Figure 20.32.

Figure 20.32 Transformers change voltages at several points in a power distribution system. Electric power is usually generated at greater than 10 kV, and transmitted long distances at voltages ranging from 120 kV to 700 kV to limit energy losses. Local power distribution to neighborhoods or industries goes through a substation and is sent short distances at voltages ranging from 5 to 13 kV. This is reduced to 120, 240, or 480 V for safety at the individual user site.

Once the power has arrived at a population or industrial center, the voltage is stepped down at a substation to between 5 and 30

kV. Finally, at individual homes or businesses, the power is stepped down again to 120, 240, or 480 V. Each step-up and step-down transformation is done with a transformer designed based on Faradays law of induction. We've come a long way since Queen Elizabeth asked Faraday what possible use could be made of electricity.

Check Your Understanding

7. What is an electric motor?
   a. An electric motor transforms electrical energy into mechanical energy.
   b. An electric motor transforms mechanical energy into electrical energy.
   c. An electric motor transforms chemical energy into mechanical energy.
   d. An electric motor transforms mechanical energy into chemical energy.

8. What happens to the torque provided by an electric motor if you double the number of coils in the motor?
   a. The torque would be doubled.
   b. The torque would be halved.
   c. The torque would be quadrupled.
   d. The torque would be tripled.

9. What is a step-up transformer?
   a. A step-up transformer decreases the current to transmit power over short distance with minimum loss.
   b. A step-up transformer increases the current to transmit power over short distance with minimum loss.
   c. A step-up transformer increases voltage to transmit power over long distance with minimum loss.
   d. A step-up transformer decreases voltage to transmit power over short distance with minimum loss.

10. What should be the ratio of the number of output coils to the number of input coil in a step-up transformer to increase the voltage fivefold?
   a. The ratio is five times.
   b. The ratio is 10 times.
   c. The ratio is 15 times.
   d. The ratio is 20 times.

20.3 Electromagnetic Induction

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Explain how a changing magnetic field produces a current in a wire
- Calculate induced electromotive force and current

Section Key Terms

emf	induction	magnetic flux

Changing Magnetic Fields

In the preceding section, we learned that a current creates a magnetic field. If nature is symmetrical, then perhaps a magnetic field can create a current. In 1831, some 12 years after the discovery that an electric current generates a magnetic field, English scientist Michael Faraday (1791–1862) and American scientist Joseph Henry (1797–1878) independently demonstrated that magnetic fields can produce currents. The basic process of generating currents with magnetic fields is called **induction**; this process is also called magnetic induction to distinguish it from charging by induction, which uses the electrostatic Coulomb force.

When Faraday discovered what is now called Faraday's law of induction, Queen Victoria asked him what possible use was electricity. "Madam," he replied, "What good is a baby?" Today, currents induced by magnetic fields are essential to our technological society. The electric generator—found in everything from automobiles to bicycles to nuclear power plants—uses magnetism to generate electric current. Other devices that use magnetism to induce currents include pickup coils in electric guitars, transformers of every size, certain microphones, airport security gates, and damping mechanisms on sensitive chemical balances.

One experiment Faraday did to demonstrate magnetic induction was to move a bar magnet through a wire coil and measure the resulting electric current through the wire. A schematic of this experiment is shown in Figure 20.33. He found that current is induced only when the magnet moves with respect to the coil. When the magnet is motionless with respect to the coil, no current is induced in the coil, as in Figure 20.33. In addition, moving the magnet in the opposite direction (compare Figure 20.33 with Figure 20.33) or reversing the poles of the magnet (compare Figure 20.33 with Figure 20.33) results in a current in the opposite direction.

Figure 20.33 Movement of a magnet relative to a coil produces electric currents as shown. The same currents are produced if the coil is moved relative to the magnet. The greater the speed, the greater the magnitude of the current, and the current is zero when there is no motion. The current produced by moving the magnet upward is in the opposite direction as the current produced by moving the magnet downward.

Virtual Physics

Faraday's Law

Click to view content (http://www.openstax.org/l/faradays-law)

Try this simulation to see how moving a magnet creates a current in a circuit. A light bulb lights up to show when current is flowing, and a voltmeter shows the voltage drop across the light bulb. Try moving the magnet through a four-turn coil and through a two-turn coil. For the same magnet speed, which coil produces a higher voltage?

GRASP CHECK

With the north pole to the left and moving the magnet from right to left, a positive voltage is produced as the magnet enters the coil. What sign voltage will be produced if the experiment is repeated with the south pole to the left?

a. The sign of voltage will change because the direction of current flow will change by moving south pole of the magnet to the left.
b. The sign of voltage will remain same because the direction of current flow will not change by moving south pole of the magnet to the left.
c. The sign of voltage will change because the magnitude of current flow will change by moving south pole of the magnet to the left.
d. The sign of voltage will remain same because the magnitude of current flow will not change by moving south pole of the magnet to the left.

Induced Electromotive Force

If a current is induced in the coil, Faraday reasoned that there must be what he called an *electromotive force* pushing the charges through the coil. This interpretation turned out to be incorrect; instead, the external source doing the work of moving the magnet adds energy to the charges in the coil. The energy added per unit charge has units of volts, so the electromotive force is actually a potential. Unfortunately, the name electromotive force stuck and with it the potential for confusing it with a real force. For this reason, we avoid the term *electromotive force* and just use the abbreviation *emf*, which has the mathematical symbol ε. The **emf** may be defined as the rate at which energy is drawn from a source per unit current flowing through a circuit. Thus, emf is the energy per unit charge *added* by a source, which contrasts with voltage, which is the energy per unit charge

released as the charges flow through a circuit.

To understand why an emf is generated in a coil due to a moving magnet, consider Figure 20.34, which shows a bar magnet moving downward with respect to a wire loop. Initially, seven magnetic field lines are going through the loop (see left-hand image). Because the magnet is moving away from the coil, only five magnetic field lines are going through the loop after a short time Δt (see right-hand image). Thus, when a change occurs in the number of magnetic field lines going through the area defined by the wire loop, an emf is induced in the wire loop. Experiments such as this show that the induced emf is proportional to the *rate of change* of the magnetic field. Mathematically, we express this as

$$\varepsilon \propto \frac{\Delta B}{\Delta t}, \tag{20.24}$$

where ΔB is the change in the magnitude in the magnetic field during time Δt and A is the area of the loop.

Figure 20.34 The bar magnet moves downward with respect to the wire loop, so that the number of magnetic field lines going through the loop decreases with time. This causes an emf to be induced in the loop, creating an electric current.

Note that magnetic field lines that lie in the plane of the wire loop do not actually pass through the loop, as shown by the left-most loop in Figure 20.35. In this figure, the arrow coming out of the loop is a vector whose magnitude is the area of the loop and whose direction is perpendicular to the plane of the loop. In Figure 20.35, as the loop is rotated from $\theta = 90°$ to $\theta = 0°$, the contribution of the magnetic field lines to the emf increases. Thus, what is important in generating an emf in the wire loop is the component of the magnetic field that is *perpendicular* to the plane of the loop, which is $B \cos \theta$.

This is analogous to a sail in the wind. Think of the conducting loop as the sail and the magnetic field as the wind. To maximize the force of the wind on the sail, the sail is oriented so that its surface vector points in the same direction as the winds, as in the right-most loop in Figure 20.35. When the sail is aligned so that its surface vector is perpendicular to the wind, as in the left-most loop in Figure 20.35, then the wind exerts no force on the sail.

Thus, taking into account the angle of the magnetic field with respect to the area, the proportionality $E \propto \Delta B/\Delta t$ becomes

$$E \propto \frac{\Delta B \cos \theta}{\Delta t}. \tag{20.25}$$

Figure 20.35 The magnetic field lies in the plane of the left-most loop, so it cannot generate an emf in this case. When the loop is rotated so that the angle of the magnetic field with the vector perpendicular to the area of the loop increases to 90° (see right-most loop), the magnetic field contributes maximally to the emf in the loop. The dots show where the magnetic field lines intersect the plane defined by the loop.

Another way to reduce the number of magnetic field lines that go through the conducting loop in Figure 20.35 is not to move the magnet but to make the loop smaller. Experiments show that changing the area of a conducting loop in a stable magnetic field induces an emf in the loop. Thus, the emf produced in a conducting loop is proportional to the rate of change of the *product* of the perpendicular magnetic field and the loop area

$$\varepsilon \propto \frac{\Delta\left[(B\cos\theta)A\right]}{\Delta t}, \tag{20.26}$$

where $B\cos\theta$ is the perpendicular magnetic field and A is the area of the loop. The product $BA\cos\theta$ is very important. It is proportional to the number of magnetic field lines that pass perpendicularly through a surface of area A. Going back to our sail analogy, it would be proportional to the force of the wind on the sail. It is called the **magnetic flux** and is represented by Φ .

$$\Phi = BA\cos\theta \tag{20.27}$$

The unit of magnetic flux is the weber (Wb), which is magnetic field per unit area, or T/m^2. The weber is also a volt second (Vs).

The induced emf is in fact proportional to the rate of change of the magnetic flux through a conducting loop.

$$\varepsilon \propto \frac{\Delta\Phi}{\Delta t} \tag{20.28}$$

Finally, for a coil made from N loops, the emf is N times stronger than for a single loop. Thus, the emf induced by a changing magnetic field in a coil of N loops is

$$\varepsilon \propto N\frac{\Delta B\cos\theta}{\Delta t}\text{A}.$$

The last question to answer before we can change the proportionality into an equation is "In what direction does the current flow?" The Russian scientist Heinrich Lenz (1804–1865) explained that the current flows in the direction that creates a magnetic field that tries to keep the flux constant in the loop. For example, consider again Figure 20.34. The motion of the bar magnet causes the number of upward-pointing magnetic field lines that go through the loop to decrease. Therefore, an emf is generated in the loop that drives a current in the direction that creates more upward-pointing magnetic field lines. By using the right-hand rule, we see that this current must flow in the direction shown in the figure. To express the fact that the induced emf acts to counter the change in the magnetic flux through a wire loop, a minus sign is introduced into the proportionality $\varepsilon \propto \Delta\Phi/\Delta t.$, which gives Faraday's law of induction.

$$\varepsilon = -N\frac{\Delta\Phi}{\Delta t} \tag{20.29}$$

Lenz's law is very important. To better understand it, consider Figure 20.36, which shows a magnet moving with respect to a wire coil and the direction of the resulting current in the coil. In the top row, the north pole of the magnet approaches the coil, so the magnetic field lines from the magnet point toward the coil. Thus, the magnetic field $\vec{B}_{\text{mag}} = B_{\text{mag}}(\hat{x})$ pointing to the right increases in the coil. According to Lenz's law, the emf produced in the coil will drive a current in the direction that creates a magnetic field $\vec{B}_{\text{coil}} = B_{\text{coil}}(-\hat{x})$ inside the coil pointing to the left. This will counter the increase in magnetic flux pointing to the right. To see which way the current must flow, point your right thumb in the desired direction of the magnetic field $\vec{B}_{\text{coil}}$, and the current will flow in the direction indicated by curling your right fingers. This is shown by the image of the right hand in the top row of Figure 20.36. Thus, the current must flow in the direction shown in Figure 4(a).

In Figure 4(b), the direction in which the magnet moves is reversed. In the coil, the right-pointing magnetic field $\vec{B}_{\text{mag}}$ due to the moving magnet decreases. Lenz's law says that, to counter this decrease, the emf will drive a current that creates an additional right-pointing magnetic field $\vec{B}_{\text{coil}}$ in the coil. Again, point your right thumb in the desired direction of the magnetic field, and the current will flow in the direction indicate by curling your right fingers (Figure 4(b)).

Finally, in Figure 4(c), the magnet is reversed so that the south pole is nearest the coil. Now the magnetic field $\vec{B}_{\text{mag}}$ points toward the magnet instead of toward the coil. As the magnet approaches the coil, it causes the left-pointing magnetic field in the coil to increase. Lenz's law tells us that the emf induced in the coil will drive a current in the direction that creates a magnetic field pointing to the right. This will counter the increasing magnetic flux pointing to the left due to the magnet. Using the right-hand rule again, as indicated in the figure, shows that the current must flow in the direction shown in Figure 4(c).

Figure 20.36 Lenz's law tells us that the magnetically induced emf will drive a current that resists the change in the magnetic flux through a circuit. This is shown in panels (a)–(c) for various magnet orientations and velocities. The right hands at right show how to apply the right-hand rule to find in which direction the induced current flows around the coil.

Virtual Physics

Faraday's Electromagnetic Lab

Click to view content (http://www.openstax.org/l/Faraday-EM-lab)

This simulation proposes several activities. For now, click on the tab Pickup Coil, which presents a bar magnet that you can move through a coil. As you do so, you can see the electrons move in the coil and a light bulb will light up or a voltmeter will indicate the voltage across a resistor. Note that the voltmeter allows you to see the sign of the voltage as you move the magnet about. You can also leave the bar magnet at rest and move the coil, although it is more difficult to observe the results.

GRASP CHECK

Orient the bar magnet with the north pole facing to the right and place the pickup coil to the right of the bar magnet. Now move the bar magnet toward the coil and observe in which way the electrons move. This is the same situation as depicted below. Does the current in the simulation flow in the same direction as shown below? Explain why or why not.

a. Yes, the current in the simulation flows as shown because the direction of current is opposite to the direction of flow of electrons.
b. No, current in the simulation flows in the opposite direction because the direction of current is same to the direction of flow of electrons.

WATCH PHYSICS

Induced Current in a Wire

This video explains how a current can be induced in a straight wire by moving it through a magnetic field. The lecturer uses the *cross product*, which a type of vector multiplication. Don't worry if you are not familiar with this, it basically combines the right-hand rule for determining the force on the charges in the wire with the equation $F = qvB \sin\theta$.

Click to view content (https://www.openstax.org/l/induced-current)

GRASP CHECK

What emf is produced across a straight wire 0.50 m long moving at a velocity of (1.5 m/s) $\hat{x}$ through a uniform magnetic field (0.30 T)$\hat{z}$? The wire lies in the $\hat{y}$-direction. Also, which end of the wire is at the higher potential—let the lower end of the wire be at $y = 0$ and the upper end at $y = 0.5$ m)?

a. 0.15 V and the lower end of the wire will be at higher potential
b. 0.15 V and the upper end of the wire will be at higher potential
c. 0.075 V and the lower end of the wire will be at higher potential
d. 0.075 V and the upper end of the wire will be at higher potential

WORKED EXAMPLE

EMF Induced in Conducing Coil by Moving Magnet

Imagine a magnetic field goes through a coil in the direction indicated in Figure 20.37. The coil diameter is 2.0 cm. If the magnetic field goes from 0.020 to 0.010 T in 34 s, what is the direction and magnitude of the induced current? Assume the coil has a resistance of 0.1 Ω.

Figure 20.37 A coil through which passes a magnetic field *B*.

STRATEGY

Use the equation $\varepsilon = -N\Delta\Phi/\Delta t$ to find the induced emf in the coil, where $\Delta t = 34\ \text{s}$. Counting the number of loops in the solenoid, we find it has 16 loops, so $N = 16$. Use the equation $\Phi = BA\cos\theta$ to calculate the magnetic flux

$$\Phi = BA\cos\theta = B\pi\left(\frac{d}{2}\right)^2, \tag{20.30}$$

where d is the diameter of the solenoid and we have used $\cos 0° = 1$. Because the area of the solenoid does not vary, the change in the magnetic of the flux through the solenoid is

$$\Delta\Phi = \Delta B\pi\left(\frac{d}{2}\right)^2. \tag{20.31}$$

Once we find the emf, we can use Ohm's law, $\varepsilon = IR$, to find the current.

Finally, Lenz's law tells us that the current should produce a magnetic field that acts to oppose the decrease in the applied magnetic field. Thus, the current should produce a magnetic field to the right.

Solution

Combining equations $\varepsilon = -N\Delta\Phi/\Delta t$ and $\Phi = BA\cos\theta$ gives

$$\varepsilon = -N\frac{\Delta\Phi}{\Delta t} = -N\frac{\Delta B\pi d^2}{4\ \Delta t}. \tag{20.32}$$

Solving Ohm's law for the current and using this result gives

$$\begin{aligned} I &= \frac{\varepsilon}{R} = -N\frac{\Delta B\pi d^2}{4R\Delta t} \\ &= -16\frac{(-0.010\ \text{T})\pi(0.020\ \text{m})^2}{4(0.10\ \Omega)(34\ \text{s})}. \\ &= 15\ \mu\text{A} \end{aligned} \tag{20.33}$$

Lenz's law tells us that the current must produce a magnetic field to the right. Thus, we point our right thumb to the right and curl our right fingers around the solenoid. The current must flow in the direction in which our fingers are pointing, so it enters at the left end of the solenoid and exits at the right end.

Discussion

Let's see if the minus sign makes sense in Faraday's law of induction. Define the direction of the magnetic field to be the positive direction. This means the change in the magnetic field is negative, as we found above. The minus sign in Faraday's law of induction negates the negative change in the magnetic field, leaving us with a positive current. Therefore, the current must flow in the direction of the magnetic field, which is what we found.

Now try defining the positive direction to be the direction opposite that of the magnetic field, that is positive is to the left in Figure 20.37. In this case, you will find a negative current. But since the positive direction is to the left, a negative current must flow to the right, which again agrees with what we found by using Lenz's law.

WORKED EXAMPLE

Magnetic Induction due to Changing Circuit Size

The circuit shown in Figure 20.38 consists of a U-shaped wire with a resistor and with the ends connected by a sliding conducting rod. The magnetic field filling the area enclosed by the circuit is constant at 0.01 T. If the rod is pulled to the right at speed $v = 0.50\ \text{m/s}$, what current is induced in the circuit and in what direction does the current flow?

Figure 20.38 A slider circuit. The magnetic field is constant and the rod is pulled to the right at speed *v*. The changing area enclosed by the circuit induces an emf in the circuit.

STRATEGY

We again use Faraday's law of induction, $E = -N\frac{\Delta\Phi}{\Delta t}$, although this time the magnetic field is constant and the area enclosed by the circuit changes. The circuit contains a single loop, so $N = 1$. The rate of change of the area is $\frac{\Delta A}{\Delta t} = v\ell$. Thus the rate of change of the magnetic flux is

$$\frac{\Delta\Phi}{\Delta t} = \frac{\Delta\,(BA\cos\theta)}{\Delta t} = B\frac{\Delta A}{\Delta t} = Bv\ell, \tag{20.34}$$

where we have used the fact that the angle θ between the area vector and the magnetic field is 0°. Once we know the emf, we can find the current by using Ohm's law. To find the direction of the current, we apply Lenz's law.

Solution

Faraday's law of induction gives

$$E = -N\frac{\Delta\Phi}{\Delta t} = -Bv\ell. \tag{20.35}$$

Solving Ohm's law for the current and using the previous result for emf gives

$$I = \frac{E}{R} = \frac{-Bv\ell}{R} = \frac{-(0.010\text{ T})\,(0.50\text{ m/s})\,(0.10\text{ m})}{20\ \Omega} = 25\ \mu\text{A}. \tag{20.36}$$

As the rod slides to the right, the magnetic flux passing through the circuit increases. Lenz's law tells us that the current induced will create a magnetic field that will counter this increase. Thus, the magnetic field created by the induced current must be into the page. Curling your right-hand fingers around the loop in the clockwise direction makes your right thumb point into the page, which is the desired direction of the magnetic field. Thus, the current must flow in the clockwise direction around the circuit.

Discussion

Is energy conserved in this circuit? An external agent must pull on the rod with sufficient force to just balance the force on a current-carrying wire in a magnetic field—recall that $F = I\ell B\sin\theta$. The rate at which this force does work on the rod should be balanced by the rate at which the circuit dissipates power. Using $F = I\ell B\sin\theta$, the force required to pull the wire at a constant speed *v* is

$$F_{\text{pull}} = I\ell B\sin\theta = I\ell B, \tag{20.37}$$

where we used the fact that the angle θ between the current and the magnetic field is 90°. Inserting our expression above for the current into this equation gives

$$F_{\text{pull}} = I\ell B = -\frac{Bv\ell}{R}(\ell B) = -\frac{B^2 v\ell^2}{R}. \tag{20.38}$$

The power contributed by the agent pulling the rod is $F_{\text{pull}}\,v$, or

$$P_{\text{pull}} = F_{\text{pull}} v = -\frac{B^2 v^2 \ell^2}{R}. \tag{20.39}$$

The power dissipated by the circuit is

$$P_{\text{dissipated}} = I^2 R = \left(\frac{-Bv\ell}{R}\right)^2 R = \frac{B^2 v^2 \ell^2}{R}. \tag{20.40}$$

We thus see that $P_{\text{pull}} + P_{\text{dissipated}} = 0$, which means that power is conserved in the system consisting of the circuit and the agent that pulls the rod. Thus, energy is conserved in this system.

Practice Problems

11. The magnetic flux through a single wire loop changes from 3.5 Wb to 1.5 Wb in 2.0 s. What emf is induced in the loop?

a. −2.0 V
b. −1.0 V
c. +1.0 V
d. +2.0 V

12. What is the emf for a 10-turn coil through which the flux changes at 10 Wb/s?

a. −100 V
b. −10 V
c. +10 V
d. +100 V

Check Your Understanding

13. Given a bar magnet, how can you induce an electric current in a wire loop?

a. An electric current is induced if a bar magnet is placed near the wire loop.
b. An electric current is induced if wire loop is wound around the bar magnet.
c. An electric current is induced if a bar magnet is moved through the wire loop.
d. An electric current is induced if a bar magnet is placed in contact with the wire loop.

14. What factors can cause an induced current in a wire loop through which a magnetic field passes?

a. Induced current can be created by changing the size of the wire loop only.
b. Induced current can be created by changing the orientation of the wire loop only.
c. Induced current can be created by changing the strength of the magnetic field only.
d. Induced current can be created by changing the strength of the magnetic field, changing the size of the wire loop, or changing the orientation of the wire loop.

KEY TERMS

Curie temperature well-defined temperature for ferromagnetic materials above which they cannot be magnetized

domain region within a magnetic material in which the magnetic poles of individual atoms are aligned

electric motor device that transforms electrical energy into mechanical energy

electromagnet device that uses electric current to make a magnetic field

electromagnetism study of electric and magnetic phenomena

emf rate at which energy is drawn from a source per unit current flowing through a circuit

ferromagnetic material such as iron, cobalt, nickel, or gadolinium that exhibits strong magnetic effects

generator device that transforms mechanical energy into electrical energy

induction rate at which energy is drawn from a source per unit current flowing through a circuit

magnetic dipole term that describes magnets because they always have two poles: north and south

magnetic field directional lines around a magnetic material that indicates the direction and magnitude of the magnetic force

magnetic flux component of the magnetic field perpendicular to the surface area through which it passes and multiplied by the area

magnetic pole part of a magnet that exerts the strongest force on other magnets or magnetic material

magnetized material that is induced to be magnetic or that is made into a permanent magnet

north pole part of a magnet that orients itself toward the geographic North Pole of Earth

permanent magnet material that retains its magnetic behavior for a long time, even when exposed to demagnetizing influences

right-hand rule rule involving curling the right-hand fingers from one vector to another; the direction in which the right thumb points is the direction of the resulting vector

solenoid uniform cylindrical coil of wire through which electric current is passed to produce a magnetic field

south pole part of a magnet that orients itself toward the geographic South Pole of Earth

transformer device that transforms voltages from one value to another

SECTION SUMMARY

20.1 Magnetic Fields, Field Lines, and Force

- All magnets have two poles: a north pole and a south pole. If the magnet is free to move, its north pole orients itself toward the geographic North Pole of Earth, and the south pole orients itself toward the geographic South Pole of Earth.
- A repulsive force occurs between the north poles of two magnets and likewise for two south poles. However, an attractive force occurs between the north pole of one magnet and the south pole of another magnet.
- A charged particle moving through a magnetic field experiences a force whose direction is determined by the right-hand rule.
- An electric current generates a magnetic field.
- Electromagnets are magnets made by passing a current through a system of wires.

20.2 Motors, Generators, and Transformers

- Electric motors contain wire loops in a magnetic field. Current is passed through the wire loops, which forces them to rotate in the magnetic field. The current is reversed every half rotation so that the torque on the loop is always in the same direction.
- Electric generators contain wire loops in a magnetic field. An external agent provides mechanical energy to force the loops to rotate in the magnetic field, which produces an AC voltage that drives an AC current through the loops.
- Transformers contain a ring made of magnetic material and, on opposite sides of the ring, two windings of wire wrap around the ring. A changing current in one wire winding creates a changing magnetic field, which is trapped in the ring and thus goes through the second winding and induces an emf in the second winding. The voltage in the second winding is proportional to the ratio of the number of loops in each winding.
- Transformers are used to step up and step down the voltage for power transmission.
- Over long distances, electric power is transmitted at high voltage to minimize the current and thereby minimize the Joule losses due to resistive heating.

20.3 Electromagnetic Induction

- **Faraday's law of induction** states that a changing magnetic flux that occurs within an area enclosed by a conducting loop induces an electric current in the loop.
- **Lenz' law** states that an induced current flows in the direction such that it opposes the change that induced it.

KEY EQUATIONS

20.1 Magnetic Fields, Field Lines, and Force

the magnitude of the force on an electric charge	$F = qvB\sin\theta$
the force on a wire carrying current	$F = I\ell B\sin\theta$
the magnitude of the magnetic field created by a long, straight current-carrying wire	$B_{\text{straightwire}} = \frac{\mu_0 I}{2\pi r}$
the magnitude of the magnetic field inside a solenoid	$B_{\text{solenoid}} = \mu_0 \frac{NI}{\ell}$

20.3 Electromagnetic Induction

magnetic flux	$\Phi = BA\cos\theta$
emf	$E = -N\frac{\Delta\Phi}{\Delta t}$

CHAPTER REVIEW

Concept Items

20.1 Magnetic Fields, Field Lines, and Force

1. If you place a small needle between the north poles of two bar magnets, will the needle become magnetized?
   a. Yes, the magnetic fields from the two north poles will point in the same directions.
   b. Yes, the magnetic fields from the two north poles will point in opposite directions.
   c. No, the magnetic fields from the two north poles will point in opposite directions.
   d. No, the magnetic fields from the two north poles will point in the same directions.

2. If you place a compass at the three points in the figure, at which point will the needle experience the greatest torque? Why?

   a. The density of the magnetic field is minimized at B, so the magnetic compass needle will experience the greatest torque at B.
   b. The density of the magnetic field is minimized at C, so the magnetic compass needle will experience the greatest torque at C.
   c. The density of the magnetic field is maximized at B, so the magnetic compass needle will experience the greatest torque at B.
   d. The density of the magnetic field is maximized at A, so the magnetic compass needle will experience the greatest torque at A.

3. In which direction do the magnetic field lines point near the south pole of a magnet?
   a. Outside the magnet the direction of magnetic field lines is towards the south pole of the magnet.
   b. Outside the magnet the direction of magnetic field lines is away from the south pole of the magnet.

20.2 Motors, Generators, and Transformers

4. Consider the angle between the area vector and the magnetic field in an electric motor. At what angles is the torque on the wire loop the greatest?
   a. 0° and 180°
   b. 45° and 135°
   c. 90° and 270°
   d. 225° and 315°

5. What is a voltage transformer?
   a. A transformer is a device that transforms current to voltage.
   b. A transformer is a device that transforms voltages from one value to another.
   c. A transformer is a device that transforms resistance of wire to voltage.

6. Why is electric power transmitted at high voltage?

a. To increase the current for the transmission
b. To reduce energy loss during transmission
c. To increase resistance during transmission
d. To reduce resistance during transmission

20.3 Electromagnetic Induction

7. Yes or no—Is an emf induced in the coil shown when it is stretched? If so, state why and give the direction of the induced current.

a. No, because induced current does not depend upon the area of the coil.
b. Yes, because area of the coil increases; the direction of the induced current is counterclockwise.
c. Yes, because area of the coil increases; the direction of the induced current is clockwise.
d. Yes, because the area of the coil does not change; the direction of the induced current is clockwise.

8. What is Lenz's law?

a. If induced current flows, its direction is such that it adds to the changes which induced it.
b. If induced current flows, its direction is such that it opposes the changes which induced it.
c. If induced current flows, its direction is always clockwise to the changes which induced it.
d. If induced current flows, its direction is always counterclockwise to the changes which induced it.

9. Explain how magnetic flux can be zero when the magnetic field is not zero.

a. If angle between magnetic field and area vector is 0°, then its sine is also zero, which means that there is zero flux.
b. If angle between magnetic field and area vector is 45°, then its sine is also zero, which means that there is zero flux.
c. If angle between magnetic field and area vector is 60°, then its cosine is also zero, which means that there is zero flux.
d. If the angle between magnetic field and area vector is 90°, then its cosine is also zero, which means that there is zero flux.

Critical Thinking Items

20.1 Magnetic Fields, Field Lines, and Force

10. True or false—It is not recommended to place credit cards with magnetic strips near permanent magnets.

a. false
b. true

11. True or false—A square magnet can have sides that alternate between north and south poles.

a. false
b. true

12. You move a compass in a circular plane around a planar magnet. The compass makes four complete revolutions. How many poles does the magnet have?

a. two poles
b. four poles
c. eight poles
d. 12 poles

20.2 Motors, Generators, and Transformers

13. How can you maximize the peak emf from a generator?

a. The peak emf from a generator can be maximized only by maximizing number of turns.
b. The peak emf from a generator can be maximized only by maximizing area of the wired loop.
c. The peak emf from a generator can be maximized only by maximizing frequency.
d. The peak emf from a generator can be maximized by maximizing number of turns, maximizing area of the wired loop or maximizing frequency.

14. Explain why power is transmitted over long distances at high voltages.

a. $P_{lost} = I_{transmitted} V_{transmitted}$, so to maximize current, the voltage must be maximized
b. $P_{transmitted} = I_{transmitted} V_{transmitted}$, so to maximize current, the voltage must be maximized
c. $P_{lost} = I_{transmitted} V_{transmitted}$, so to minimize current, the voltage must be maximized
d. $P_{transmitted} = I_{transmitted} V_{transmitted}$, so to minimize current, the voltage must be maximized

20.3 Electromagnetic Induction

15. To obtain power from the current in the wire of your vacuum cleaner, you place a loop of wire near it to obtain an induced emf. How do you place and orient the loop?

a. A loop of wire should be placed nearest to the vacuum cleaner wire to maximize the magnetic flux through the loop.
b. A loop of wire should be placed farthest to the vacuum cleaner wire to maximize the magnetic flux through the loop.

c. A loop of wire should be placed perpendicular to the vacuum cleaner wire to maximize the magnetic flux through the loop.
d. A loop of wire should be placed at angle greater than 90° to the vacuum cleaner wire to maximize the magnetic flux through the loop.

16. A magneto is a device that creates a spark across a gap by creating a large voltage across the gap. To do this, the device spins a magnet very quickly in front of a wire coil, with the ends of the wires forming the gap. Explain how this creates a sufficiently large voltage to produce a spark.
a. The electric field in the coil increases rapidly due to spinning of magnet which creates an emf in the coil that is proportional to the rate of change of the magnetic flux.
b. The magnetic field in the coil changes rapidly due to spinning of magnet which creates an emf in the coil that is proportional to the rate of change of the magnetic flux.

17. If you drop a copper tube over a bar magnet with its north pole up, is a current induced in the copper tube? If so, in what direction? Consider when the copper tube is approaching the bar magnet.
a. Yes, the induced current will be produced in the clockwise direction when viewed from above.
b. No, the induced current will not be produced.

Problems

20.1 Magnetic Fields, Field Lines, and Force

18. A straight wire segment carries 0.25 A. What length would it need to be to exert a 4.0-mN force on a magnet that produces a uniform magnetic field of 0.015 T that is perpendicular to the wire?
a. 0.55 m
b. 1.10 m
c. 2.20 m
d. 4.40 m

20.3 Electromagnetic Induction

19. What is the current in a wire loop of resistance 10 Ω through which the magnetic flux changes from zero to 10 Wb in 1.0 s?
a. −100 A
b. −2.0 A
c. −1.0 A
d. +1.0 A

20. An emf is induced by rotating a 1,000 turn, 20.0 cm diameter coil in Earth's 5.00×10^{-5} T magnetic field. What average emf is induced, given the plane of the coil is originally perpendicular to Earth's field and is rotated to be parallel to the field in 10.0 ms?
a. -1.6×10^{-4} V
b. $+1.6 \times 10^{-4}$ V
c. $+1.6 \times 10^{-1}$ V
d. -1.6×10^{-1} V

Performance Task

20.2 Motors, Generators, and Transformers

21. Your family takes a trip to Cuba, and rents an old car to drive into the countryside to see the sights. Unfortunately, the next morning you find yourself deep in the countryside and the car won't start because the battery is too weak. Wanting to jump-start the car, you open the hood and find that you can't tell which battery terminal is positive and which is negative. However, you do have a bar magnet with the north and south poles labeled and you manage to find a short wire. How do you use these to determine which terminal is which? For starters, how do you determine the direction of a magnetic field around a current-carrying wire? And in which direction will the force be on another magnet placed in this field? Do you need to worry about the sign of the mobile charge carriers in the wire?

TEST PREP

Multiple Choice

20.1 Magnetic Fields, Field Lines, and Force

22. For a magnet, a domain refers to _____.
a. the region between the poles of the magnet
b. the space around the magnet that is affected by the magnetic field
c. the region within the magnet in which the magnetic poles of individual atoms are aligned
d. the region from which the magnetic material is mined

23. In the region just outside the south pole of a magnet, the magnetic field lines _____.
a. point away from the south pole
b. go around the south pole
c. are less concentrated than at the north pole

d. point toward the south pole

24. Which equation gives the force for a charge moving through a magnetic field?
a. $F = qvB \sin\theta$
b. $F = I\ell B \sin\theta$
c. $F = I\ell B$
d. $F = qvB$

25. Can magnetic field lines cross each other? Explain why or why not.
a. Yes, magnetic field lines can cross each other because that point of intersection indicates two possible directions of magnetic field, which is possible.
b. No, magnetic field lines cannot cross each other because that point of intersection indicates two possible directions of magnetic field, which is not possible.

26. True or false—If a magnet shatters into many small pieces, all the pieces will have north and south poles
a. true
b. false

20.2 Motors, Generators, and Transformers

27. An electrical generator _____.
a. is a generator powered by electricity
b. must be turned by hand
c. converts other sources of power into electrical power
d. uses magnetism to create electrons

28. A step-up transformer increases the
a. voltage from power lines for use in homes
b. current from the power lines for use in homes
c. current from the electrical generator for transmission along power lines
d. voltage from the electrical power plant for transmission along power lines

29. What would be the effect on the torque of an electric motor of doubling the width of the current loop in the motor?
a. Torque remains the same.
b. Torque is doubled.
c. Torque is quadrupled.
d. Torque is halved.

30. Why are the coils of a transformer wrapped around a loop of ferrous material?
a. The magnetic field from the source coil is trapped and also increased in strength.
b. The magnetic field from the source coil is dispersed and also increased in strength.
c. The magnetic field from the source coil is trapped and also decreased in strength.
d. Magnetic field from the source coil is dispersed and also decreased in strength.

20.3 Electromagnetic Induction

31. What does *emf* stand for?
a. electromotive force
b. electro motion force
c. electromagnetic factor
d. electronic magnetic factor

32. Which formula gives magnetic flux?
a. $\frac{\mu_0 I}{2\pi r}$
b. $qvB \sin\theta$
c. $-N\frac{\Delta\Phi}{\Delta t}$
d. $BA\cos\theta$

33. What is the relationship between the number of coils in a solenoid and the emf induced in it by a change in the magnetic flux through the solenoid?
a. The induced emf is inversely proportional to the number of coils in a solenoid.
b. The induced emf is directly proportional to the number of coils in a solenoid.
c. The induced emf is inversely proportional to the square of the number of coils in a solenoid.
d. The induced emf is proportional to square of the number of coils in a solenoid.

34. True or false—If you drop a bar magnet through a copper tube, it induces an electric current in the tube.
a. false
b. true

Short Answer

20.1 Magnetic Fields, Field Lines, and Force

35. Given a bar magnet, a needle, a cork, and a bowl full of water, describe how to make a compass.
a. Magnetize the needle by holding it perpendicular to a bar magnet's north pole and pierce the cork along its longitudinal axis by the needle and place the needle-cork combination in the water. The needle now orients itself along the magnetic field lines of Earth.
b. Magnetize the needle by holding it perpendicular to a bar magnet's north pole and pierce the cork along its longitudinal axis by the needle and place the needle-cork combination in the water. The needle now orients itself perpendicular to the

magnetic field lines of Earth.

c. Magnetize the needle by holding its axis parallel to the axis of a bar magnet and pierce the cork along its longitudinal axis by the needle and place the needle-cork combination in the water. The needle now orients itself along the magnetic field lines of Earth.

d. Magnetize the needle by holding its axis parallel to the axis of a bar magnet and pierce the cork along its longitudinal axis by the needle and place the needle-cork combination in the water. The needle now orients itself perpendicular to the magnetic field lines of Earth.

36. Give two differences between electric field lines and magnetic field lines.

a. Electric field lines begin and end on opposite charges and the electric force on a charge is in the direction of field, while magnetic fields form a loop and the magnetic force on a charge is perpendicular to the field.

b. Electric field lines form a loop and the electric force on a charge is in the direction of field, while magnetic fields begin and end on opposite charge and the magnetic force on a charge is perpendicular to the field.

c. Electric field lines begin and end on opposite charges and the electric force on a charge is in the perpendicular direction of field, while magnetic fields form a loop and the magnetic force on a charge is in the direction of the field.

d. Electric field lines form a loop and the electric force on a charge is in the perpendicular direction of field, while magnetic fields begin and end on opposite charge and the magnetic force on a charge is in the direction of the field.

37. To produce a magnetic field of 0.0020 T, what current is required in a 500-turn solenoid that is 25 cm long?

a. 0.80 A

b. 1.60 A

c. 80 A

d. 160 A

38. You magnetize a needle by aligning it along the axis of a bar magnet and just outside the north pole of the magnet. Will the point of the needle that was closest to the bar magnet then be attracted to or repelled from the south pole of another magnet?

a. The needle will magnetize and the point of needle kept closer to the north pole will act as a south pole. Hence, it will repel the south pole of other magnet.

b. The needle will magnetize and the point of needle kept closer to the north pole will act as a south pole. Hence, it will attract the south pole of other magnet.

c. The needle will magnetize and the point of a needle kept closer to the north pole will act as a north pole. Hence, it will repel the south pole of the other magnet.

d. The needle will magnetize and the point of needle kept closer to the north pole will act as a north pole. Hence, it will attract the south pole of other magnet.

39. Using four solenoids of the same size, describe how to orient them and in which direction the current should flow to make a magnet with two opposite-facing north poles and two opposite-facing south poles.

40. How far from a straight wire carrying 0.45 A is the magnetic field strength 0.040 T?

a. 0.23 μm
b. 0.72 μm
c. 2.3 μm
d. 7.2 μm

20.2 Motors, Generators, and Transformers

41. A laminated-coil transformer has a wire coiled 12 times around one of its sides. How many coils should you wrap around the opposite side to get a voltage output that is one half of the input voltage? Explain.

a. six output coils because the ratio of output to input voltage is the same as the ratio of number of output coils to input coils
b. 12 output coils because the ratio of output to input voltage is the same as the ratio of number of output coils to input coils
c. 24 output coils because the ratio of output to input voltage is half the ratio of the number of output coils to input coils
d. 36 output coils because the ratio of output to input voltage is three times the ratio of the number of output coils to input coils

42. Explain why long-distance electrical power lines are designed to carry very high voltages.

a. $P_{transmitted} = I_{transmitted>}{}^2 R_{wire}$ and $P_{lost} = I_{transmitted}$ $V_{transmitted}$, so V must be low to make the current transmitted as high as possible.
b. $P_{transmitted} = I_{transmitted>}{}^2 R_{wire}$ and $P_{lost} = I_{lost}$ V_{lost}, so V must be low to make the current transmitted as high as possible.
c. $P_{transmitted} = I_{transmitted>}{}^2 R_{wire}$ and $P_{lost} = I_{transmitted}$ $V_{transmitted}$, so V must be high to make the current transmitted as low as possible
d. $P_{lost} = I_{transmitted}{}^2 R_{wire}$ and $P_{transmitted} = I_{transmitted}$ $V_{transmitted}$, so V must be high to make the current transmitted as low as possible.

43. How is the output emf of a generator affected if you double the frequency of rotation of its coil?

a. The output emf will be doubled.
b. The output emf will be halved.
c. The output emf will be quadrupled.
d. The output emf will be tripled.

44. In a hydroelectric dam, what is used to power the electrical generators that provide electric power? Explain.

a. The electric potential energy of stored water is used to produce emf with the help of a turbine.
b. The electric potential energy of stored water is used to produce resistance with the help of a turbine.
c. Gravitational potential energy of stored water is used to produce resistance with the help of a turbine.
d. Gravitational potential energy of stored water is used to produce emf with the help of a turbine.

20.3 Electromagnetic Induction

45. A uniform magnetic field is perpendicular to the plane of a wire loop. If the loop accelerates in the direction of the field, will a current be induced in the loop? Explain why or why not.

a. No, because magnetic flux through the loop remains constant.
b. No, because magnetic flux through the loop changes continuously.
c. Yes, because magnetic flux through the loop remains constant.
d. Yes, because magnetic flux through the loop changes continuously.

46. The plane of a square wire circuit with side 4.0 cm long is at an angle of 45° with respect to a uniform magnetic field of 0.25 T. The wires have a resistance per unit length of 0.2. If the field drops to zero in 2.5 s, what magnitude current is induced in the square circuit?

a. 35 μA
b. 87.5 μA
c. 3.5 mA
d. 35 A

47. Yes or no—If a bar magnet moves through a wire loop as shown in the figure, is a current induced in the loop? Explain why or why not.

a. No, because the net magnetic field passing through the loop is zero.
b. No, because the net magnetic field passing through

the loop is nonzero.

c. Yes, because the net magnetic field passing through the loop is zero.

d. Yes, because the net magnetic field line passing through the loop is nonzero.

48. What is the magnetic flux through an equilateral triangle with side 60 cm long and whose plane makes a 60° angle with a uniform magnetic field of 0.33 T?

a. 0.045 Wb

b. 0.09 Wb

c. 0.405 Wb

d. 4.5 Wb

Extended Response

20.1 Magnetic Fields, Field Lines, and Force

49. Summarize the properties of magnets.

a. A magnet can attract metals like iron, nickel, etc., but cannot attract nonmetals like piece of plastic or wood, etc. If free to rotate, an elongated magnet will orient itself so that its north pole will face the magnetic south pole of Earth.

b. A magnet can attract metals like iron, nickel, etc., but cannot attract nonmetals like piece of plastic or wood, etc. If free to rotate, an elongated magnet will orient itself so that its north pole will face the magnetic north pole of Earth.

c. A magnet can attract metals like iron, nickel, etc., and nonmetals like piece of plastic or wood, etc. If free to rotate, an elongated magnet will orient itself so that its north pole will face the magnetic south pole of Earth.

d. A magnet can attract metals like iron, nickel, etc., and nonmetals like piece of plastic or wood, etc. If free to rotate, an elongated magnet will orient itself so that its north pole will face the magnetic north pole of Earth.

50. The magnetic field shown in the figure is formed by current flowing in two rings that intersect the page at the dots. Current flows into the page at the dots with crosses (right side) and out of the page at the dots with points (left side).

Where is the field strength the greatest and in what direction do the magnetic field lines point?

a. The magnetic field strength is greatest where the magnetic field lines are less dense; magnetic field lines points up the page.

b. The magnetic field strength is greatest where the magnetic field lines are most dense; magnetic field lines points up the page.

c. The magnetic field strength is greatest where the magnetic field lines are most dense; magnetic field lines points down the page.

d. The magnetic field strength is greatest where the magnetic field lines are less dense; magnetic field lines points down the page.

51. The forces shown below are exerted on an electron as it moves through the magnetic field. In each case, what direction does the electron move?

a. (a) left to right, (b) out of the page, (c) upwards

b. (a) left to right, (b) into the page, (c) downwards

c. (a) right to left, (b) out of the page, (c) upwards

d. (a) right to left, (b) into the page, (c) downwards

20.2 Motors, Generators, and Transformers

52. Explain why increasing the frequency of rotation of the coils in an electrical generator increases the output emf.

a. The induced emf is proportional to the rate of change of magnetic flux with respect to distance.

b. The induced emf is inversely proportional to the rate of change of magnetic flux with respect to distance.

c. The induced emf is inversely proportional to the rate of change of magnetic flux with respect to time.

d. The induced emf is proportional to the rate of change of magnetic flux with respect to time.

53. Your friend tells you that power lines must carry a

maximum current because $P = I^2R$, where R is the resistance of the transmission line. What do you tell her?

a. $P_{transmitted} = I_{transmitted}{}^2 R_{wire}$ and $P_{lost} = I_{transmitted}$ $V_{transmitted}$, so I must be high to reduce power lost due to transmission.

b. $P_{lost} = I_{transmitted}{}^2 R_{wire}$ and $P_{lost} = I_{transmitted}$ $V_{transmitted}$, so I must be high to reduce power lost due to transmission.

c. $P_{transmitted} = I_{transmitted}{}^2 R_{wire}$ and $P_{lost} = I_{transmitted}$ $V_{transmitted}$, so I must be low to reduce power lost due to transmission.

d. $P_{lost} = I_{transmitted}{}^2 R_{wire}$ and $P_{lost} = I_{transmitted}$ $V_{transmitted}$, so I must be low to reduce power lost due to transmission.

20.3 Electromagnetic Induction

54. When you insert a copper ring between the poles of two bar magnets as shown in the figure, do the magnets exert an attractive or repulsive force on the ring? Explain your reasoning.

a. Magnets exert an attractive force, because magnetic field due to induced current is repulsed by the magnetic field of the magnets.

b. Magnets exert an attractive force, because magnetic field due to induced current is attracted by the magnetic field of the magnets.

c. Magnets exert a repulsive force, because magnetic field due to induced current is repulsed by the magnetic field of the magnets.

d. Magnets exert a repulsive force, because magnetic field due to induced current is attracted by the magnetic field of the magnets.

55. The figure shows a uniform magnetic field passing through a closed wire circuit. The wire circuit rotates at an angular frequency of about the axis shown by the dotted line in the figure.

What is an expression for the magnetic flux through the circuit as a function of time?

a. expression for the magnetic flux through the circuit $\Phi(t) = BA\cos\omega t$

b. expression for the magnetic flux through the circuit $\Phi(t) = \sqrt{2}BA\cos\omega t$

c. expression for the magnetic flux through the circuit $\Phi(t) = \sqrt{3}BA\cos\omega t$

d. expression for the magnetic flux through the circuit $\Phi(t) = 2BA\cos\omega t$

CHAPTER 21
The Quantum Nature of Light

Figure 21.1 In Lewis Carroll's classic text Alice's Adventures in Wonderland, Alice follows a rabbit down a hole into a land of curiosity. While many of her interactions in Wonderland are of surprising consequence, they follow a certain inherent logic. (credit: modification of work by John Tenniel, Wikimedia Commons)

Chapter Outline

21.1 Planck and Quantum Nature of Light

21.2 Einstein and the Photoelectric Effect

21.3 The Dual Nature of Light

INTRODUCTION At first glance, the quantum nature of light can be a strange and bewildering concept. Between light acting as discrete chunks, massless particles providing momenta, and fundamental particles behaving like waves, it may often seem like something out of Alice in Wonderland.

For many, the study of this branch of physics can be as enthralling as Lewis Carroll's classic novel. Recalling the works of legendary characters and brilliant scientists such as Einstein, Planck, and Compton, the study of light's quantum nature will provide you an interesting tale of how a clever interpretation of some small details led to the most important discoveries of the past 150 years. From the electronics revolution of the twentieth century to our future progress in solar energy and space exploration, the quantum nature of light should yield a rabbit hole of curious consequence, within which lie some of the most fascinating truths of our time.

21.1 Planck and Quantum Nature of Light

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Describe blackbody radiation
- Define quantum states and their relationship to modern physics
- Calculate the quantum energy of lights
- Explain how photon energies vary across divisions of the electromagnetic spectrum

Section Key Terms

blackbody	quantized	quantum	ultraviolet catastrophe

Blackbodies

Our first story of curious significance begins with a T-shirt. You are likely aware that wearing a tight black T-shirt outside on a hot day provides a significantly less comfortable experience than wearing a white shirt. Black shirts, as well as all other black objects, will absorb and re-emit a significantly greater amount of radiation from the sun. This shirt is a good approximation of what is called a **blackbody**.

A perfect blackbody is one that absorbs and re-emits all radiated energy that is incident upon it. Imagine wearing a tight shirt that did this! This phenomenon is often modeled with quite a different scenario. Imagine carving a small hole in an oven that can be heated to very high temperatures. As the temperature of this container gets hotter and hotter, the radiation out of this dark hole would increase as well, re-emitting all energy provided it by the increased temperature. The hole may even begin to glow in different colors as the temperature is increased. Like a burner on your stove, the hole would glow red, then orange, then blue, as the temperature is increased. In time, the hole would continue to glow but the light would be invisible to our eyes. This container is a good model of a perfect blackbody.

It is the analysis of blackbodies that led to one of the most consequential discoveries of the twentieth century. Take a moment to carefully examine Figure 21.2. What relationships exist? What trends can you see? The more time you spend interpreting this figure, the closer you will be to understanding quantum physics!

Figure 21.2 Graphs of blackbody radiation (from an ideal radiator) at three different radiator temperatures. The intensity or rate of radiation emission increases dramatically with temperature, and the peak of the spectrum shifts toward the visible and ultraviolet parts of the spectrum. The shape of the spectrum cannot be described with classical physics.

TIPS FOR SUCCESS

When encountering a new graph, it is best to try to interpret the graph before you read about it. Doing this will make the following text more meaningful and will help to remind yourself of some of the key concepts within the section.

Understanding Blackbody Graphs

Figure 21.2 is a plot of radiation intensity against radiated wavelength. In other words, it shows how the intensity of radiated light changes when a blackbody is heated to a particular temperature.

It may help to just follow the bottom-most red line labeled 3,000 K, red hot. The graph shows that when a blackbody acquires a temperature of 3,000 K, it radiates energy across the electromagnetic spectrum. However, the energy is most intensely emitted at a wavelength of approximately 1000 nm. This is in the infrared portion of the electromagnetic spectrum. While a body at this temperature would appear *red-hot* to our eyes, it would truly appear 'infrared-hot' if we were able to see the entire spectrum.

A few other important notes regarding Figure 21.2:

- As temperature increases, the total amount of energy radiated increases. This is shown by examining the area underneath each line.
- Regardless of temperature, all red lines on the graph undergo a consistent pattern. While electromagnetic radiation is emitted throughout the spectrum, the intensity of this radiation peaks at one particular wavelength.
- As the temperature changes, the wavelength of greatest radiation intensity changes. At 4,000 K, the radiation is most intense in the yellow-green portion of the spectrum. At 6,000 K, the blackbody would radiate *white hot,* due to intense radiation throughout the visible portion of the electromagnetic spectrum. Remember that white light is the emission of all visible colors simultaneously.
- As the temperature increases, the frequency of light providing the greatest intensity increases as well. Recall the equation $v = f\lambda$. Because the speed of light is constant, frequency and wavelength are inversely related. This is verified by the leftward movement of the three red lines as temperature is increased.

While in science it is important to categorize observations, theorizing as to why the observations exist is crucial to scientific advancement. Why doesn't a blackbody emit radiation evenly across all wavelengths? Why does the temperature of the body change the peak wavelength that is radiated? Why does an increase in temperature cause the peak wavelength emitted to decrease? It is questions like these that drove significant research at the turn of the twentieth century. And within the context of these questions, Max Planck discovered something of tremendous importance.

Planck's Revolution

The prevailing theory at the time of Max Planck's discovery was that intensity and frequency were related by the equation $I = \frac{2kT}{\lambda^2}$. This equation, derived from classical physics and using wave phenomena, infers that as wavelength increases, the intensity of energy provided will decrease with an inverse-squared relationship. This relationship is graphed in Figure 21.3 and shows a troubling trend. For starters, it should be apparent that the graph from this equation does not match the blackbody graphs found experimentally. Additionally, it shows that for an object of any temperature, there should be an infinite amount of energy quickly emitted in the shortest wavelengths. When theory and experimental results clash, it is important to re-evaluate both models. The disconnect between theory and reality was termed the **ultraviolet catastrophe**.

Figure 21.3 The graph above shows the true spectral measurements by a blackbody against those predicted by the classical theory at the time. The discord between the predicted classical theory line and the actual results is known as the ultraviolet catastrophe.

Due to concerns over the ultraviolet catastrophe, Max Planck began to question whether another factor impacted the relationship between intensity and wavelength. This factor, he posited, should affect the probability that short wavelength light would be emitted. Should this factor reduce the probability of short wavelength light, it would cause the radiance curve to not progress infinitely as in the classical theory, but would instead cause the curve to precipitate back downward as is shown in the 5,000 K, 4,000 K, and 3,000 K temperature lines of the graph in Figure 21.3. Planck noted that this factor, whatever it may be, must also be dependent on temperature, as the intensity decreases at lower and lower wavelengths as the temperature increases.

The determination of this *probability factor* was a groundbreaking discovery in physics, yielding insight not just into light but also into energy and matter itself. It would be the basis for Planck's 1918 Nobel Prize in Physics and would result in the transition of physics from classical to modern understanding. In an attempt to determine the cause of the *probability factor,* Max Planck constructed a new theory. This theory, which created the branch of physics called quantum mechanics, speculated that the energy radiated by the blackbody could exist only in specific numerical, or **quantum**, states. This theory is described by the equation $E = nhf$, where n is any nonnegative integer (0, 1, 2, 3, ...) and h is Planck's constant, given by $h = 6.626 \times 10^{-34}\,\text{J} \cdot \text{s}$, and f is frequency.

Through this equation, Planck's probability factor can be more clearly understood. Each frequency of light provides a specific quantized amount of energy. Low frequency light, associated with longer wavelengths would provide a smaller amount of energy, while high frequency light, associated with shorter wavelengths, would provide a larger amount of energy. For specified temperatures with specific total energies, it makes sense that more low frequency light would be radiated than high frequency light. To a degree, the relationship is like pouring coins through a funnel. More of the smaller pennies would be able to pass through the funnel than the larger quarters. In other words, because the value of the coin is somewhat related to the size of the coin, the probability of a quarter passing through the funnel is reduced!

Furthermore, an increase in temperature would signify the presence of higher energy. As a result, the greater amount of total blackbody energy would allow for more of the high frequency, short wavelength, energies to be radiated. This permits the peak of the blackbody curve to drift leftward as the temperature increases, as it does from the 3,000 K to 4,000 K to 5,000 K values. Furthering our coin analogy, consider a wider funnel. This funnel would permit more quarters to pass through and allow for a reduction in concern about the *probability factor*.

In summary, it is the interplay between the predicted classical model and the quantum probability that creates the curve depicted in Figure 21.3. Just as quarters have a higher currency denomination than pennies, higher frequencies come with larger

amounts of energy. However, just as the probability of a quarter passing through a fixed diameter funnel is reduced, so is the probability of a high frequency light existing in a fixed temperature object. As is often the case in physics, it is the balancing of multiple incredible ideas that finally allows for better understanding.

Quantization

It may be helpful at this point to further consider the idea of quantum states. Atoms, molecules, and fundamental electron and proton charges are all examples of physical entities that are **quantized**—that is, they appear only in certain discrete values and do not have every conceivable value. On the macroscopic scale, this is not a revolutionary concept. A standing wave on a string allows only particular harmonics described by integers. Going up and down a hill using discrete stair steps causes your potential energy to take on discrete values as you move from step to step. Furthermore, we cannot have a fraction of an atom, or part of an electron's charge, or 14.33 cents. Rather, everything is built of integral multiples of these substructures.

That said, to discover quantum states within a phenomenon that science had always considered continuous would certainly be surprising. When Max Planck was able to use quantization to correctly describe the experimentally known shape of the blackbody spectrum, it was the first indication that energy was quantized on a small scale as well. This discovery earned Planck the Nobel Prize in Physics in 1918 and was such a revolutionary departure from classical physics that Planck himself was reluctant to accept his own idea. The general acceptance of Planck's energy quantization was greatly enhanced by Einstein's explanation of the photoelectric effect (discussed in the next section), which took energy quantization a step further.

Figure 21.4 The German physicist Max Planck had a major influence on the early development of quantum mechanics, being the first to recognize that energy is sometimes quantized. Planck also made important contributions to special relativity and classical physics. (credit: Library of Congress, Prints and Photographs Division, Wikimedia Commons)

WORKED EXAMPLE

How Many Photons per Second Does a Typical Light Bulb Produce?

Assuming that 10 percent of a 100-W light bulb's energy output is in the visible range (typical for incandescent bulbs) with an average wavelength of 580 nm, calculate the number of visible photons emitted per second.

Strategy

The number of visible photons per second is directly related to the amount of energy emitted each second, also known as the bulb's power. By determining the bulb's power, the energy emitted each second can be found. Since the power is given in watts, which is joules per second, the energy will be in joules. By comparing this to the amount of energy associated with each photon, the number of photons emitted each second can be determined.

Solution

The power in visible light production is 10.0 percent of 100 W, or 10.0 J/s. The energy of the average visible photon is found by substituting the given average wavelength into the formula

$$E = nhf = \frac{nhc}{\lambda}.$$

By rearranging the above formula to determine energy per photon, this produces

$$E/n = \frac{(6.63 \times 10^{-34}\,\text{J}\cdot\text{s})(3.00 \times 10^{8}\,\text{m/s})}{580 \times 10^{-9}\,\text{m}} = 3.43 \times 10^{-19}\,\text{J/photon}. \quad (21.1)$$

The number of visible photons per second is thus

$$\frac{\text{photons}}{\text{sec}} = \frac{10.0\text{J/s}}{3.43 \times 10^{-19}\text{J/photon}} = 2.92 \times 10^{19}\text{photons/s}.$$

Discussion

This incredible number of photons per second is verification that individual photons are insignificant in ordinary human experience. However, it is also a verification of our everyday experience—on the macroscopic scale, photons are so small that quantization becomes essentially continuous.

WORKED EXAMPLE

How does Photon Energy Change with Various Portions of the EM Spectrum?

Refer to the Graphs of Blackbody Radiation shown in the first figure in this section. Compare the energy necessary to radiate one photon of infrared light and one photon of visible light.

Strategy

To determine the energy radiated, it is necessary to use the equation $E = nhf$. It is also necessary to find a representative frequency for infrared light and visible light.

Solution

According to the first figure in this section, one representative wavelength for infrared light is 2000 nm (2.000×10^{-6} m). The associated frequency of an infrared light is

$$f = \frac{c}{\lambda} = \frac{3.00 \times 10^{8}\text{m/s}}{2.000 \times 10^{-6}\text{m}} = 1.50 \times 10^{14}\text{Hz}. \tag{21.2}$$

Using the equation $E = nhf$, the energy associated with one photon of representative infrared light is

$$\frac{E}{n} = h \cdot f = \left(6.63 \times 10^{-34}\text{J} \cdot \text{s}\right)\left(1.50 \times 10^{14}\text{Hz}\right) = 9.95 \times 10^{-20}\frac{\text{J}}{\text{photon}}. \tag{21.3}$$

The same process above can be used to determine the energy associated with one photon of representative visible light. According to the first figure in this section, one representative wavelength for visible light is 500 nm.

$$f = \frac{c}{\lambda} = \frac{3.00 \times 10^{8}\text{m/s}}{5.00 \times 10^{-7}\text{m}} = 6.00 \times 10^{14}\text{Hz}. \tag{21.4}$$

$$\frac{E}{n} = h \cdot f = \left(6.63 \times 10^{-34}\text{J} \cdot \text{s}\right)\left(6.00 \times 10^{14}\text{Hz}\right) = 3.98 \times 10^{-19}\frac{\text{J}}{\text{photon}}. \tag{21.5}$$

Discussion

This example verifies that as the wavelength of light decreases, the quantum energy increases. This explains why a fire burning with a blue flame is considered more dangerous than a fire with a red flame. Each photon of short-wavelength blue light emitted carries a greater amount of energy than a long-wavelength red light. This example also helps explain the differences in the 3,000 K, 4,000 K, and 6,000 K lines shown in the first figure in this section. As the temperature is increased, more energy is available for a greater number of short-wavelength photons to be emitted.

Practice Problems

1. An AM radio station broadcasts at a frequency of 1,530 kHz . What is the energy in Joules of a photon emitted from this station?
   a. 10.1×10^{-26} J
   b. 1.01×10^{-28} J
   c. 1.01×10^{-29} J
   d. 1.01×10^{-27} J

2. A photon travels with energy of 1.0 eV. What type of EM radiation is this photon?
   a. visible radiation

b. microwave radiation
c. infrared radiation
d. ultraviolet radiation

Check Your Understanding

3. Do reflective or absorptive surfaces more closely model a perfect blackbody?
   a. reflective surfaces
   b. absorptive surfaces

4. A black T-shirt is a good model of a blackbody. However, it is not perfect. What prevents a black T-shirt from being considered a perfect blackbody?
   a. The T-shirt reflects some light.
   b. The T-shirt absorbs all incident light.
   c. The T-shirt re-emits all the incident light.
   d. The T-shirt does not reflect light.

5. What is the mathematical relationship linking the energy of a photon to its frequency?
   a. $E = \frac{hf}{n}$
   b. $E = \frac{nh}{f}$
   c. $E = \frac{nf}{h}$
   d. $E = nhf$

6. Why do we not notice quantization of photons in everyday experience?
   a. because the size of each photon is very large
   b. because the mass of each photon is so small
   c. because the energy provided by photons is very large
   d. because the energy provided by photons is very small

7. Two flames are observed on a stove. One is red while the other is blue. Which flame is hotter?
   a. The red flame is hotter because red light has lower frequency.
   b. The red flame is hotter because red light has higher frequency.
   c. The blue flame is hotter because blue light has lower frequency.
   d. The blue flame is hotter because blue light has higher frequency.

8. Your pupils dilate when visible light intensity is reduced. Does wearing sunglasses that lack UV blockers increase or decrease the UV hazard to your eyes? Explain.
   a. Increase, because more high-energy UV photons can enter the eye.
   b. Increase, because less high-energy UV photons can enter the eye.
   c. Decrease, because more high-energy UV photons can enter the eye.
   d. Decrease, because less high-energy UV photons can enter the eye.

9. The temperature of a blackbody radiator is increased. What will happen to the most intense wavelength of light emitted as this increase occurs?
   a. The wavelength of the most intense radiation will vary randomly.
   b. The wavelength of the most intense radiation will increase.
   c. The wavelength of the most intense radiation will remain unchanged.
   d. The wavelength of the most intense radiation will decrease.

21.2 Einstein and the Photoelectric Effect

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Describe Einstein's explanation of the photoelectric effect
- Describe how the photoelectric effect could not be explained by classical physics
- Calculate the energy of a photoelectron under given conditions
- Describe use of the photoelectric effect in biological applications, photoelectric devices and movie soundtracks

Section Key Terms

electric eye	photoelectric effect	photoelectron	photon

The Photoelectric Effect

Teacher Support
[EL]Ask the students what they think the term *photoelectric* means. How does the term relate to its definition?

When light strikes certain materials, it can eject electrons from them. This is called the **photoelectric effect**, meaning that light (*photo*) produces electricity. One common use of the photoelectric effect is in light meters, such as those that adjust the automatic iris in various types of cameras. Another use is in solar cells, as you probably have in your calculator or have seen on a rooftop or a roadside sign. These make use of the photoelectric effect to convert light into electricity for running different devices.

Figure 21.5 The photoelectric effect can be observed by allowing light to fall on the metal plate in this evacuated tube. Electrons ejected by the light are collected on the collector wire and measured as a current. A retarding voltage between the collector wire and plate can then be adjusted so as to determine the energy of the ejected electrons. (credit: P. P. Urone)

Revolutionary Properties of the Photoelectric Effect

When Max Planck theorized that energy was quantized in a blackbody radiator, it is unlikely that he would have recognized just how revolutionary his idea was. Using tools similar to the light meter in Figure 21.5, it would take a scientist of Albert Einstein's stature to fully discover the implications of Max Planck's radical concept.

Through careful observations of the photoelectric effect, Albert Einstein realized that there were several characteristics that could be explained only if *EM radiation is itself quantized*. While these characteristics will be explained a bit later in this section, you can already begin to appreciate why Einstein's idea is very important. It means that the apparently continuous stream of energy in an EM wave is actually not a continuous stream at all. In fact, the EM wave itself is actually composed of tiny quantum packets of energy called **photons**.

In equation form, Einstein found the energy of a photon or **photoelectron** to be

$$E = h\mathrm{f},$$

where E is the energy of a photon of frequency f and h is Planck's constant. A beam from a flashlight, which to this point had been considered a wave, instead could now be viewed as a series of photons, each providing a specific amount of energy see Figure 21.6. Furthermore, the amount of energy within each individual photon is based upon its individual frequency, as

dictated by $E = hf$. As a result, the total amount of energy provided by the beam could now be viewed as the sum of all frequency-dependent photon energies added together.

Figure 21.6 An EM wave of frequency *f* is composed of photons, or individual quanta of EM radiation. The energy of each photon is $E = hf$, where *h* is Planck's constant and *f* is the frequency of the EM radiation. Higher intensity means more photons per unit area per second. The flashlight emits large numbers of photons of many different frequencies, hence others have energy $E' = hf'$, and so on.

Just as with Planck's blackbody radiation, Einstein's concept of the photon could take hold in the scientific community only if it could succeed where classical physics failed. The photoelectric effect would be a key to demonstrating Einstein's brilliance.

Consider the following five properties of the photoelectric effect. All of these properties are consistent with the idea that individual photons of EM radiation are absorbed by individual electrons in a material, with the electron gaining the photon's energy. Some of these properties are inconsistent with the idea that EM radiation is a simple wave. For simplicity, let us consider what happens with monochromatic EM radiation in which all photons have the same energy *hf*.

Figure 21.7 Incident radiation strikes a clean metal surface, ejecting multiple electrons from it. The manner in which the frequency and intensity of the incoming radiation affect the ejected electrons strongly suggests that electromagnetic radiation is quantized. This event, called the photoelectric effect, is strong evidence for the existence of photons.

1. If we vary the frequency of the EM radiation falling on a clean metal surface, we find the following: For a given material, there is a threshold frequency f_o for the EM radiation below which no electrons are ejected, regardless of intensity. Using the photon model, the explanation for this is clear. Individual photons interact with individual electrons. Thus if the energy of an individual photon is too low to break an electron away, no electrons will be ejected. However, if EM radiation were a simple wave, sufficient energy could be obtained simply by increasing the intensity.
2. *Once EM radiation falls on a material, electrons are ejected without delay.* As soon as an individual photon of sufficiently high frequency is absorbed by an individual electron, the electron is ejected. If the EM radiation were a simple wave, several minutes would be required for sufficient energy to be deposited at the metal surface in order to eject an electron.
3. The number of electrons ejected per unit time is proportional to the intensity of the EM radiation and to no other characteristic. High-intensity EM radiation consists of large numbers of photons per unit area, with all photons having the same characteristic energy, *hf*. The increased number of photons per unit area results in an increased number of electrons per unit area ejected.
4. If we vary the intensity of the EM radiation and measure the energy of ejected electrons, we find the following: *The maximum kinetic energy of ejected electrons is independent of the intensity of the EM radiation.* Instead, as noted in point 3 above, increased intensity results in more electrons of the same energy being ejected. If EM radiation were a simple wave, a higher intensity could transfer more energy, and higher-energy electrons would be ejected.
5. The kinetic energy KE of an ejected electron equals the photon energy minus the binding energy BE of the electron in the

specific material. An individual photon can give all of its energy to an electron. The photon's energy is partly used to break the electron away from the material. The remainder goes into the ejected electron's kinetic energy. In equation form, this is given by

$$KE_e = hf - BE, \tag{21.6}$$

where KE_e is the maximum kinetic energy of the ejected electron, hf is the photon's energy, and BE is the binding energy of the electron to the particular material. This equation explains the properties of the photoelectric effect quantitatively and demonstrates that BE is the minimum amount of energy necessary to eject an electron. If the energy supplied is less than BE, the electron cannot be ejected. The binding energy can also be written as $BE = hf_0$, where f_0 is the threshold frequency for the particular material. Figure 21.8 shows a graph of maximum KE_e versus the frequency of incident EM radiation falling on a particular material.

Figure 21.8 A graph of the kinetic energy of an ejected electron, KE_e, versus the frequency of EM radiation impinging on a certain material. There is a threshold frequency below which no electrons are ejected, because the individual photon interacting with an individual electron has insufficient energy to break it away. Above the threshold energy, KE_e increases linearly with f, consistent with $KE_e = hf - BE$. The slope of this line is h, so the data can be used to determine Planck's constant experimentally.

TIPS FOR SUCCESS

The following five pieces of information can be difficult to follow without some organization. It may be useful to create a table of expected results of each of the five properties, with one column showing the classical wave model result and one column showing the modern photon model result.
The table may look something like Table 21.1

	Classical Wave Model	Modern Photon Model
Threshold Frequency		
Electron Ejection Delay		
Intensity of EM Radiation		
Speed of Ejected Electrons		
Relationship between Kinetic Energy and Binding Energy		

Table 21.1 Table of Expected Results

Virtual Physics

Photoelectric Effect

Click to view content (http://www.openstax.org/l/28photoelectric)

In this demonstration, see how light knocks electrons off a metal target, and recreate the experiment that spawned the field of quantum mechanics.

GRASP CHECK

In the circuit provided, what are the three ways to increase the current?

a. decrease the intensity, decrease the frequency, alter the target
b. decrease the intensity, decrease the frequency, don't alter the target
c. increase the intensity, increase the frequency, alter the target
d. increase the intensity, increase the frequency, alter the target

WORKED EXAMPLE

Photon Energy and the Photoelectric Effect: A Violet Light

(a) What is the energy in joules and electron volts of a photon of 420-nm violet light? (b) What is the maximum kinetic energy of electrons ejected from calcium by 420 nm violet light, given that the binding energy of electrons for calcium metal is 2.71 eV?

Strategy

To solve part (a), note that the energy of a photon is given by $E = hf$. For part (b), once the energy of the photon is calculated, it is a straightforward application of $KE_e = hf - BE$ to find the ejected electron's maximum kinetic energy, since BE is given.

Solution for (a)

Photon energy is given by

$$E = hf.$$

Since we are given the wavelength rather than the frequency, we solve the familiar relationship $c = f\lambda$ for the frequency, yielding

$$f = \frac{c}{\lambda} \tag{21.7}$$

Combining these two equations gives the useful relationship

$$E = \frac{hc}{\lambda}. \tag{21.8}$$

Now substituting known values yields

$$E = \frac{\left(6.63 \times 10^{-34}\text{J} \cdot \text{s}\right)\left(3.00 \times 10^{8} m/s\right)}{4.20 \times 10^{-7} m} = 4.74 \times 10^{-19}\text{J}. \tag{21.9}$$

Converting to eV, the energy of the photon is

$$E = (4.74 \times 10^{-19}\text{J} \cdot \text{s})\frac{1 \text{ eV}}{1.60 \times 10^{-19}\text{J}} = 2.96 \text{ eV}. \tag{21.10}$$

Solution for (b)

Finding the kinetic energy of the ejected electron is now a simple application of the equation $KE_e = hf - BE$. Substituting the photon energy and binding energy yields

$$KE_e = hf - BE = 2.96 \text{ eV} - 2.71 \text{ eV} = 0.25 \text{ eV}. \tag{21.11}$$

Discussion

The energy of this 420 nm photon of violet light is a tiny fraction of a joule, and so it is no wonder that a single photon would be difficult for us to sense directly—humans are more attuned to energies on the order of joules. But looking at the energy in electron volts, we can see that this photon has enough energy to affect atoms and molecules. A DNA molecule can be broken with about 1 eV of energy, for example, and typical atomic and molecular energies are on the order of eV, so that the photon in this example could have biological effects, such as sunburn. The ejected electron has rather low energy, and it would not travel far,

except in a vacuum. The electron would be stopped by a retarding potential of only 0.26 eV, a slightly larger KE than calculated above. In fact, if the photon wavelength were longer and its energy less than 2.71 eV, then the formula would give a negative kinetic energy, an impossibility. This simply means that the 420 nm photons with their 2.96 eV energy are not much above the frequency threshold. You can see for yourself that the threshold wavelength is 458 nm (blue light). This means that if calcium metal were used in a light meter, the meter would be insensitive to wavelengths longer than those of blue light. Such a light meter would be completely insensitive to red light, for example.

Practice Problems

10. What is the longest-wavelength EM radiation that can eject a photoelectron from silver, given that the bonding energy is 4.73 eV ? Is this radiation in the visible range?

a. 2.63×10^{-7} m; No, the radiation is in microwave region.
b. 2.63×10^{-7} m; No, the radiation is in visible region.
c. 2.63×10^{-7} m; No, the radiation is in infrared region.
d. 2.63×10^{-7} m; No, the radiation is in ultraviolet region.

11. What is the maximum kinetic energy in eV of electrons ejected from sodium metal by 450-nm EM radiation, given that the binding energy is 2.28 eV?

a. 0.48 V
b. 0.82 eV
c. 1.21 eV
d. 0.48 eV

Technological Applications of the Photoelectric Effect

While Einstein's understanding of the photoelectric effect was a transformative discovery in the early 1900s, its presence is ubiquitous today. If you have watched streetlights turn on automatically in response to the setting sun, stopped elevator doors from closing simply by putting your hands between them, or turned on a water faucet by sliding your hands near it, you are familiar with the **electric eye**, a name given to a group of devices that use the photoelectric effect for detection.

All these devices rely on photoconductive cells. These cells are activated when light is absorbed by a semi-conductive material, knocking off a free electron. When this happens, an electron void is left behind, which attracts a nearby electron. The movement of this electron, and the resultant chain of electron movements, produces a current. If electron ejection continues, further holes are created, thereby increasing the electrical conductivity of the cell. This current can turn switches on and off and activate various familiar mechanisms.

One such mechanism takes place where you may not expect it. Next time you are at the movie theater, pay close attention to the sound coming out of the speakers. This sound is actually created using the photoelectric effect! The audiotape in the projector booth is a transparent piece of film of varying width. This film is fed between a photocell and a bright light produced by an exciter lamp. As the transparent portion of the film varies in width, the amount of light that strikes the photocell varies as well. As a result, the current in the photoconductive circuit changes with the width of the filmstrip. This changing current is converted to a changing frequency, which creates the soundtrack commonly heard in the theater.

WORK IN PHYSICS

Solar Energy Physicist

According to the U.S. Department of Energy, Earth receives enough sunlight each hour to power the entire globe for a year. While converting all of this energy is impossible, the job of the solar energy physicist is to explore and improve upon solar energy conversion technologies so that we may harness more of this abundant resource.

The field of solar energy is not a new one. For over half a century, satellites and spacecraft have utilized photovoltaic cells to create current and power their operations. As time has gone on, scientists have worked to adapt this process so that it may be used in homes, businesses, and full-scale power stations using solar cells like the one shown in Figure 21.9.

Figure 21.9 A solar cell is an example of a photovoltaic cell. As light strikes the cell, the cell absorbs the energy of the photons. If this energy exceeds the binding energy of the electrons, then electrons will be forced to move in the cell, thereby producing a current. This current may be used for a variety of purposes. (credit: U.S. Department of Energy)

Solar energy is converted to electrical energy in one of two manners: direct transfer through photovoltaic cells or thermal conversion through the use of a CSP, concentrating solar power, system. Unlike electric eyes, which trip a mechanism when current is lost, photovoltaic cells utilize semiconductors to directly transfer the electrons released through the photoelectric effect into a directed current. The energy from this current can then be converted for storage, or immediately used in an electric process. A CSP system is an indirect method of energy conversion. In this process, light from the Sun is channeled using parabolic mirrors. The light from these mirrors strikes a thermally conductive material, which then heats a pool of water. This water, in turn, is converted to steam, which turns a turbine and creates electricity. While indirect, this method has long been the traditional means of large-scale power generation.

There are, of course, limitations to the efficacy of solar power. Cloud cover, nightfall, and incident angle strike at high altitudes are all factors that directly influence the amount of light energy available. Additionally, the creation of photovoltaic cells requires rare-earth minerals that can be difficult to obtain. However, the major role of a solar energy physicist is to find ways to improve the efficiency of the solar energy conversion process. Currently, this is done by experimenting with new semi conductive materials, by refining current energy transfer methods, and by determining new ways of incorporating solar structures into the current power grid.

Additionally, many solar physicists are looking into ways to allow for increased solar use in impoverished, more remote locations. Because solar energy conversion does not require a connection to a large-scale power grid, research into thinner, more mobile materials will permit remote cultures to use solar cells to convert sunlight collected during the day into stored energy that can then be used at night.

Regardless of the application, solar energy physicists are an important part of the future in responsible energy growth. While a doctoral degree is often necessary for advanced research applications, a bachelor's or master's degree in a related science or engineering field is typically enough to gain access into the industry. Computer skills are very important for energy modeling, including knowledge of CAD software for design purposes. In addition, the ability to collaborate and communicate with others is critical to becoming a solar energy physicist.

GRASP CHECK

What role does the photoelectric effect play in the research of a solar energy physicist?

a. The understanding of photoelectric effect allows the physicist to understand the generation of light energy when using photovoltaic cells.
b. The understanding of photoelectric effect allows the physicist to understand the generation of electrical energy when using photovoltaic cells.
c. The understanding of photoelectric effect allows the physicist to understand the generation of electromagnetic energy when using photovoltaic cells.
d. The understanding of photoelectric effect allows the physicist to understand the generation of magnetic energy when using photovoltaic cells.

Check Your Understanding

12. How did Einstein's model of photons change the view of a beam of energy leaving a flashlight?
   a. A beam of light energy is now considered a continual stream of wave energy, not photons.
   b. A beam of light energy is now considered a collection of photons, each carrying its own individual energy.

13. True or false—Visible light is the only type of electromagnetic radiation that can cause the photoelectric effect.
   a. false
   b. true

14. Is the photoelectric effect a direct consequence of the wave character of EM radiation or the particle character of EM radiation?
   a. The photoelectric effect is a direct consequence of the particle nature of EM radiation.
   b. The photoelectric effect is a direct consequence of the wave nature of EM radiation.
   c. The photoelectric effect is a direct consequence of both the wave and particle nature of EM radiation.
   d. The photoelectric effect is a direct consequence of neither the wave nor the particle nature of EM radiation.

15. Which aspects of the photoelectric effect can only be explained using photons?
   a. aspects 1, 2, and 3
   b. aspects 1, 2, and 4
   c. aspects 1, 2, 4 and 5
   d. aspects 1, 2, 3, 4 and 5

16. In a photovoltaic cell, what energy transformation takes place?
   a. Solar energy transforms into electric energy.
   b. Solar energy transforms into mechanical energy.
   c. Solar energy transforms into thermal energy.
   d. In a photovoltaic cell, thermal energy transforms into electric energy.

17. True or false—A current is created in a photoconductive cell, even if only one electron is expelled from a photon strike.
   a. false
   b. true

18. What is a photon and how is it different from other fundamental particles?
   a. A photon is a quantum packet of energy; it has infinite mass.
   b. A photon is a quantum packet of energy; it is massless.
   c. A photon is a fundamental particle of an atom; it has infinite mass.
   d. A photon is a fundamental particle of an atom; it is massless.

21.3 The Dual Nature of Light

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Describe the Compton effect
- Calculate the momentum of a photon
- Explain how photon momentum is used in solar sails
- Explain the particle-wave duality of light

Section Key Terms

Compton effect	particle-wave duality	photon momentum

Photon Momentum

Do photons abide by the fundamental properties of physics? Can packets of electromagnetic energy possibly follow the same rules as a ping-pong ball or an electron? Although strange to consider, the answer to both questions is yes.

Despite the odd nature of photons, scientists prior to Einstein had long suspected that the fundamental particle of

electromagnetic radiation shared properties with our more macroscopic particles. This is no clearer than when considering the photoelectric effect, where photons knock electrons out of a substance. While it is strange to think of a massless particle exhibiting momentum, it is now a well-established fact within the scientific community. Figure 21.10 shows macroscopic evidence of **photon momentum**.

Figure 21.10 The tails of the Hale-Bopp comet point away from the Sun, evidence that light has momentum. Dust emanating from the body of the comet forms this tail. Particles of dust are pushed away from the Sun by light reflecting from them. The blue, ionized gas tail is also produced by photons interacting with atoms in the comet material. (credit: Geoff Chester, U.S. Navy, via Wikimedia Commons)

Figure 21.10 shows a comet with two prominent tails. Comet tails are composed of gases and dust evaporated from the body of the comet and ionized gas. What most people do not know about the tails is that they always point *away* from the Sun rather than trailing behind the comet. This can be seen in the diagram.

Why would this be the case? The evidence indicates that the dust particles of the comet are forced away from the Sun when photons strike them. Evidently, photons carry momentum in the direction of their motion away from the Sun, and some of this momentum is transferred to dust particles in collisions. The blue tail is caused by the solar wind, a stream of plasma consisting primarily of protons and electrons evaporating from the corona of the Sun.

Momentum, The Compton Effect, and Solar Sails

Momentum is conserved in quantum mechanics, just as it is in relativity and classical physics. Some of the earliest direct experimental evidence of this came from the scattering of X-ray photons by electrons in substances, a phenomenon discovered by American physicist Arthur H. Compton (1892–1962). Around 1923, Compton observed that X-rays reflecting from materials had decreased energy and correctly interpreted this as being due to the scattering of the X-ray photons by electrons. This phenomenon could be handled as a collision between two particles—a photon and an electron at rest in the material. After careful observation, it was found that both energy and momentum were conserved in the collision. See Figure 21.11. For the discovery of this conserved scattering, now known as the **Compton effect**, Arthur Compton was awarded the Nobel Prize in 1929.

Shortly after the discovery of Compton scattering, the value of the photon momentum, $\mathbf{p} = \frac{h}{\lambda}$,

was determined by Louis de Broglie. In this equation, called the de Broglie relation, h represents Planck's constant and λ is the photon wavelength.

Figure 21.11 The Compton effect is the name given to the scattering of a photon by an electron. Energy and momentum are conserved, resulting in a reduction of both for the scattered photon.

We can see that photon momentum is small, since $\mathbf{p} = h/\lambda$. and h is very small. It is for this reason that we do not ordinarily observe photon momentum. Our mirrors do not recoil when light reflects from them, except perhaps in cartoons. Compton saw the effects of photon momentum because he was observing X-rays, which have a small wavelength and a relatively large momentum, interacting with the lightest of particles, the electron.

WORKED EXAMPLE

Electron and Photon Momentum Compared

(a) Calculate the momentum of a visible photon that has a wavelength of 500 nm. (b) Find the velocity of an electron having the same momentum. (c) What is the energy of the electron, and how does it compare with the energy of the photon?

Strategy

Finding the photon momentum is a straightforward application of its definition: $\mathbf{p} = h/\lambda$. If we find the photon momentum is small, we can assume that an electron with the same momentum will be nonrelativistic, making it easy to find its velocity and kinetic energy from the classical formulas.

Solution for (a)

Photon momentum is given by the de Broglie relation.

$$\mathbf{p} = \frac{h}{\lambda} \tag{21.12}$$

Entering the given photon wavelength yields

$$\mathbf{p} = \frac{6.63 \times 10^{-34} J \cdot s}{5.00 \times 10^{-7} \text{m}} = 1.33 \times 10^{-27} \text{kg} \cdot \text{m/s}. \tag{21.13}$$

Solution for (b)

Since this momentum is indeed small, we will use the classical expression $p = mv$ to find the velocity of an electron with this momentum. Solving for v and using the known value for the mass of an electron gives

$$v = \frac{\text{p}}{m} = \frac{1.33 \times 10^{-27} \text{kg} \cdot \text{m/s}}{9.11 \times 10^{-31} \text{kg}} = 1{,}459.9 \text{ m/s} \approx 1{,}460 \text{ m/s}. \tag{21.14}$$

Solution for (c)

The electron has kinetic energy, which is classically given by

$$KE_e = \frac{1}{2}mv^2. \tag{21.15}$$

Thus,

$$KE_e = \frac{1}{2}(9.11 \times 10^{-31} \text{kg})(1{,}456 \text{m/s})^2 = 9.64 \times 10^{-25} \text{J}. \tag{21.16}$$

Converting this to eV by multiplying by $\frac{(1 \text{ eV})}{(1.602 \times 10^{-19} \text{J})}$ yields

$$KE_e = 6.02 \times 10^{-6} \text{eV}. \tag{21.17}$$

The photon energy E is

$$E = \frac{hc}{\lambda} = \frac{\left(6.63 \times 10^{-34} \text{J} \cdot \text{s}\right)\left(3.00 \times 10^{8} \text{ m/s}\right)}{5.00 \times 10^{-7} \text{m}} = 3.98 \times 10^{-19} \text{ J} = 2.48 \text{ eV}, \tag{21.18}$$

which is about five orders of magnitude greater.

Discussion

Even in huge numbers, the total momentum that photons carry is small. An electron that carries the same momentum as a 500-nm photon will have a 1,460 m/s velocity, which is clearly nonrelativistic. This is borne out by the experimental observation that it takes far less energy to give an electron the same momentum as a photon. That said, for high-energy photons interacting with small masses, photon momentum may be significant. Even on a large scale, photon momentum can have an effect if there

are enough of them and if there is nothing to prevent the slow recoil of matter. Comet tails are one example, but there are also proposals to build space sails that use huge low-mass mirrors (made of aluminized Mylar) to reflect sunlight. In the vacuum of space, the mirrors would gradually recoil and could actually accelerate spacecraft within the solar system. See the following figure.

TIPS FOR SUCCESS

When determining energies in particle physics, it is more sensible to use the unit eV instead of Joules. Using eV will help you to recognize differences in magnitude more easily and will make calculations simpler. Also, eV is used by scientists to describe the binding energy of particles and their rest mass, so using eV will eliminate the need to convert energy quantities. Finally, eV is a convenient unit when linking electromagnetic forces to particle physics, as one eV is the amount energy given to an electron when placed in a field of 1-V potential difference.

Practice Problems

19. Find the momentum of a 4.00-cm wavelength microwave photon.

a. 0.83×10^{-32} kg · m/s
b. 1.66×10^{-34} kg · m/s
c. 0.83×10^{-34} kg · m/s
d. 1.66×10^{-32} kg · m/s

20. Calculate the wavelength of a photon that has the same momentum of a proton moving at 1.00 percent of the speed of light.

a. 2.43×10^{-10} m
b. 2.43×10^{-12} m
c. 1.32×10^{-15} m
d. 1.32×10^{-13} m

Figure 21.12 (a) Space sails have been proposed that use the momentum of sunlight reflecting from gigantic low-mass sails to propel spacecraft about the solar system. A Russian test model of this (the Cosmos 1) was launched in 2005, but did not make it into orbit due to a rocket failure. (b) A U.S. version of this, labeled LightSail-1, is scheduled for trial launches in 2016. It will have a 40 m2 sail. (credit: Kim Newton/NASA)

LINKS TO PHYSICS

LightSail-1 Project

"Provide ships or sails adapted to the heavenly breezes, and there will be some who will brave even that void."

— Johannes Kepler (in a letter to Galileo Galilei in 1608)

Figure 21.13 NASA's NanoSail-D, a precursor to LightSail-1, with its sails deployed. The Planetary Society will be launching LightSail-1 in early 2016. (credit: NASA/MSFC/D, Wikimedia Commons)

Traversing the Solar System using nothing but the Sun's power has long been a fantasy of scientists and science fiction writers alike. Though physicists like Compton, Einstein, and Planck all provided evidence of light's propulsive capacity, it is only recently that the technology has become available to truly put these visions into motion. In 2016, by sending a lightweight satellite into space, the LightSail-1 project is designed to do just that.

A citizen-funded project headed by the Planetary Society, the 5.45-million-dollar LightSail-1 project is set to launch two crafts into orbit around the Earth. Each craft is equipped with a 32-square-meter solar sail prepared to unfurl once a rocket has launched it to an appropriate altitude. The sails are made of large mirrors, each a quarter of the thickness of a trash bag, which will receive an impulse from the Sun's reflecting photons. Each time the Sun's photon strikes the craft's reflective surface and bounces off, it will provide a momentum to the sail much greater than if the photon were simply absorbed.

Attached to three tiny satellites called CubeSats, whose combined volume is no larger than a loaf of bread, the received momentum from the Sun's photons should be enough to record a substantial increase in orbital speed. The intent of the LightSail-1 mission is to prove that the technology behind photon momentum travel is sound and can be done cheaply. A test flight in May 2015 showed that the craft's Mylar sails could unfurl on command. With another successful result in 2016, the Planetary Society will be planning future versions of the craft with the hopes of eventually achieving interplanetary satellite travel. Though a few centuries premature, Kepler's fantastic vision may not be that far away.

If eventually set into interplanetary launch, what will be the effect of continual photon bombardment on the motion of a craft similar to LightSail-1?

a. It will result in continual acceleration of the craft.
b. It will first accelerate and then decelerate the craft.
c. It will first decelerate and then accelerate the craft.
d. It will result in the craft moving at constant velocity.

Particle-Wave Duality

We have long known that EM radiation is like a wave, capable of interference and diffraction. We now see that light can also be modeled as particles—massless photons of discrete energy and momentum. We call this twofold nature the **particle-wave duality**, meaning that EM radiation has properties of both particles and waves. This may seem contradictory, since we ordinarily deal with large objects that never act like both waves and particles. An ocean wave, for example, looks nothing like a grain of sand. However, this so-called duality is simply a term for properties of the photon analogous to phenomena we can observe directly, on a macroscopic scale. See Figure 21.14. If this term seems strange, it is because we do not ordinarily observe details on the quantum level directly, and our observations yield either particle-like *or* wave-like properties, but never both simultaneously.

Figure 21.14 (a) The interference pattern for light through a double slit is a wave property understood by analogy to water waves. (b) The properties of photons having quantized energy and momentum and acting as a concentrated unit are understood by analogy to macroscopic particles.

Since we have a particle-wave duality for photons, and since we have seen connections between photons and matter in that both have momentum, it is reasonable to ask whether there is a particle-wave duality for matter as well. If the EM radiation we once thought to be a pure wave has particle properties, is it possible that matter has wave properties? The answer, strangely, is yes. The consequences of this are tremendous, as particle-wave duality has been a constant source of scientific wonder during the twentieth and twenty-first centuries.

Check Your Understanding

21. What fundamental physics properties were found to be conserved in Compton scattering?
- a. energy and wavelength
- b. energy and momentum
- c. mass and energy
- d. energy and angle

22. Why do classical or relativistic momentum equations not work in explaining the conservation of momentum that occurs in Compton scattering?
- a. because neither classical nor relativistic momentum equations utilize mass as a variable in their equations
- b. because relativistic momentum equations utilize mass as a variable in their formulas but classical momentum equations do not
- c. because classical momentum equations utilize mass as a variable in their formulas but relativistic momentum equations do not
- d. because both classical and relativistic momentum equations utilize mass as a variable in their formulas

23. If solar sails were constructed with more massive materials, how would this influence their effectiveness?
- a. The effect of the momentum would increase due to the decreased inertia of the sails.
- b. The effect of the momentum would reduce due to the decreased inertia of the sails.
- c. The effect of the momentum would increase due to the increased inertia of the sails.
- d. The effect of the momentum would be reduced due to the increased inertia of the sails.

24. True or false—It is possible to propel a solar sail craft using just particles within the solar wind.
- a. true
- b. false

25. True or false—Photon momentum more directly supports the wave model of light.
- a. false
- b. true

26. True or false—wave-particle duality exists for objects on the macroscopic scale.
   a. false
   b. true

27. What type of electromagnetic radiation was used in Compton scattering?
   a. visible light
   b. ultraviolet radiation
   c. radio waves
   d. X-rays

KEY TERMS

blackbody object that absorbs all radiated energy that strikes it and also emits energy across all wavelengths of the electromagnetic spectrum

Compton effect phenomenon whereby X-rays scattered from materials have decreased energy

electric eye group of devices that use the photoelectric effect for detection

particle-wave duality property of behaving like either a particle or a wave; the term for the phenomenon that all particles have wave-like characteristics and waves have particle-like characteristics

photoelectric effect phenomenon whereby some materials eject electrons when exposed to light

photoelectron electron that has been ejected from a material by a photon of light

photon a quantum, or particle, of electromagnetic radiation

photon momentum amount of momentum of a photon, calculated by $\mathbf{p} = \frac{h}{\lambda}$

quantized the fact that certain physical entities exist only with particular discrete values and not every conceivable value

quantum discrete packet or bundle of a physical entity such as energy

ultraviolet catastrophe misconception that blackbodies would radiate high frequency energy at a much higher rate than energy radiated at lower frequencies

SECTION SUMMARY

21.1 Planck and Quantum Nature of Light

- A blackbody will radiate energy across all wavelengths of the electromagnetic spectrum.
- Radiation of a blackbody will peak at a particular wavelength, dependent on the temperature of the blackbody.
- Analysis of blackbody radiation led to the field of quantum mechanics, which states that radiated energy can only exist in discrete quantum states.

21.2 Einstein and the Photoelectric Effect

- The photoelectric effect is the process in which EM radiation ejects electrons from a material.
- Einstein proposed photons to be quanta of EM radiation having energy $E = hf$, where f is the frequency of the radiation.
- All EM radiation is composed of photons. As Einstein explained, all characteristics of the photoelectric effect are due to the interaction of individual photons with individual electrons.
- The maximum kinetic energy KE_e of ejected electrons (photoelectrons) is given by $KE_e = hf - BE$, where hf is the photon energy and BE is the binding energy (or work function) of the electron in the particular material.

21.3 The Dual Nature of Light

- Compton scattering provided evidence that photon-electron interactions abide by the principles of conservation of momentum and conservation of energy.
- The momentum of individual photons, quantified by $\mathbf{p} = \frac{h}{\lambda}$, can be used to explain observations of comets and may lead to future space technologies.
- Electromagnetic waves and matter have both wave-like and particle-like properties. This phenomenon is defined as particle-wave duality.

KEY EQUATIONS

21.1 Planck and Quantum Nature of Light

quantum energy $E = nhf$

21.2 Einstein and the Photoelectric Effect

energy of a photon $E = hf$

maximum kinetic energy of a photoelectron $KE_e = hf - BE$

binding energy of an electron $BE = hf_O$

21.3 The Dual Nature of Light

momentum of a photon (deBroglie relation) $\mathbf{p} = \frac{h}{\lambda}$

CHAPTER REVIEW

Concept Items

21.1 Planck and Quantum Nature of Light

1. What aspect of the blackbody spectrum forced Planck to propose quantization of energy levels in atoms and molecules?
   a. Radiation occurs at a particular frequency that does not change with the energy supplied.
   b. Certain radiation occurs at a particular frequency that changes with the energy supplied.
   c. Maximum radiation would occur at a particular frequency that does not change with the energy supplied.
   d. Maximum radiation would occur at a particular frequency that changes with the energy supplied.

2. Two lasers shine red light at 650 nm. One laser is twice as bright as the other. Explain this difference using photons and photon energy.
   a. The brighter laser emits twice the number of photons and more energy per photon.
   b. The brighter laser emits twice the number of photons and less energy per photon.
   c. Both lasers emit equal numbers of photons and equivalent amounts of energy per photon.
   d. The brighter laser emits twice the number of photons but both lasers emit equivalent amounts of energy per photon.

3. Consider four stars in the night sky: red, yellow, orange, and blue. The photons of which star will carry the greatest amount of energy?
   a. blue
   b. orange
   c. red
   d. yellow

4. A lightbulb is wired to a variable resistor. What will happen to the color spectrum emitted by the bulb as the resistance of the circuit is increased?
   a. The bulb will emit greener light.
   b. The bulb will emit bluer light.
   c. The bulb will emit more ultraviolet light.
   d. The bulb will emit redder light.

21.2 Einstein and the Photoelectric Effect

5. Light is projected onto a semi-conductive surface. However, no electrons are ejected. What will happen when the light intensity is increased?
   a. An increase in light intensity decreases the number of photons. However, no electrons are ejected.
   b. Increase in light intensity increases the number of photons, so electrons with higher kinetic energy are ejected.
   c. An increase in light intensity increases the number of photons, so electrons will be ejected.
   d. An increase in light intensity increases the number of photons. However, no electrons are ejected.

6. True or false—The concept of a work function (or binding energy) is permissible under the classical wave model.
   a. false
   b. true

7. Can a single microwave photon cause cell damage?
   a. No, there is not enough energy associated with a single microwave photon to result in cell damage.
   b. No, there is zero energy associated with a single microwave photon, so it does not result in cell damage.
   c. Yes, a single microwave photon causes cell damage because it does not have high energy.
   d. Yes, a single microwave photon causes cell damage because it has enough energy.

21.3 The Dual Nature of Light

8. Why don't we feel the momentum of sunlight when we are on the beach?
   a. The momentum of a singular photon is incredibly small.
   b. The momentum is not felt because very few photons strike us at any time, and not all have momentum.
   c. The momentum of a singular photon is large, but very few photons strike us at any time.
   d. A large number of photons strike us at any time, and so their combined momentum is incredibly large.

9. If a beam of helium atoms is projected through two slits and onto a screen, will an interference pattern emerge?
   a. No, an interference pattern will not emerge because helium atoms will strike a variety of locations on the screen.
   b. No, an interference pattern will not emerge because helium atoms will strike at certain locations on the screen.
   c. Yes, an interference pattern will emerge because helium atoms will strike a variety of locations on the screen.
   d. Yes, an interference pattern will emerge because helium atoms will strike at certain locations on the screen.

Critical Thinking Items

21.1 Planck and Quantum Nature of Light

10. Explain why the frequency of a blackbody does not double when the temperature is doubled.

a. Frequency is inversely proportional to temperature.
b. Frequency is directly proportional to temperature.
c. Frequency is directly proportional to the square of temperature.
d. Frequency is directly proportional to the fourth power of temperature.

11. Why does the intensity shown in the blackbody radiation graph decrease after its peak frequency is achieved?

a. Because after reaching the peak frequency, the photons created at a particular frequency are too many for energy intensity to continue to decrease.
b. Because after reaching the peak frequency, the photons created at a particular frequency are too few for energy intensity to continue to decrease.
c. Because after reaching the peak frequency, the photons created at a particular frequency are too many for energy intensity to continue to increase.
d. Because after reaching the peak frequency, the photons created at a particular frequency are too few for energy intensity to continue to increase.

12. Shortly after the introduction of photography, it was found that photographic emulsions were more sensitive to blue and violet light than they were to red light. Explain why this was the case.

a. Blue-violet light contains greater amount of energy than red light.
b. Blue-violet light contains lower amount of energy than red light.
c. Both blue-violet light and red light have the same frequency but contain different amounts of energy.
d. Blue-violet light frequency is lower than the frequency of red light.

13. Why is it assumed that a perfect absorber of light (like a blackbody) must also be a perfect emitter of light?

a. To achieve electrostatic equilibrium with its surroundings
b. To achieve thermal equilibrium with its surroundings
c. To achieve mechanical equilibrium with its surroundings
d. To achieve chemical equilibrium with its surroundings

21.2 Einstein and the Photoelectric Effect

14. Light is projected onto a semi-conductive surface. If the intensity is held constant but the frequency of light is increased, what will happen?

a. As frequency is increased, electrons will stop being ejected from the surface.
b. As frequency is increased, electrons will begin to be ejected from the surface.
c. As frequency is increased, it will have no effect on the electrons being ejected as the intensity is the same.
d. As frequency is increased, the rate at which the electrons are being ejected will increase.

15. Why is it important to consider what material to use when designing a light meter? Consider the worked example from Section 21-2 for assistance.

a. A light meter should contain material that responds only to high frequency light.
b. A light meter should contain material that responds to low frequency light.
c. A light meter should contain material that has high binding energy.
d. A light meter should contain a material that does not show any photoelectric effect.

16. Why does overexposure to UV light often result in sunburn when overexposure to visible light does not? This is why you can get burnt even on a cloudy day.

a. UV light carries less energy than visible light and can penetrate our body.
b. UV light carries more energy than visible light, so it cannot break bonds at the cellular level.
c. UV light carries more energy than visible light and can break bonds at the cellular level.
d. UV light carries less energy than visible light and cannot penetrate the human body.

17. If you pick up and shake a piece of metal that has electrons in it free to move as a current, no electrons fall out. Yet if you heat the metal, electrons can be boiled off. Explain both of these facts as they relate to the amount and distribution of energy involved with shaking the

object as compared with heating it.

a. Thermal energy is added to the metal at a much higher rate than energy added due to shaking.
b. Thermal energy is added to the metal at a much lower rate than energy added due to shaking.
c. If the thermal energy added is below the binding energy of the electrons, they may be *boiled off.*
d. If the mechanical energy added is below the binding energy of the electrons, they may be *boiled off.*

21.3 The Dual Nature of Light

18. In many macroscopic collisions, a significant amount of kinetic energy is converted to thermal energy. Explain why this is not a concern for Compton scattering.

a. Because, photons and electrons do not exist on the molecular level, all energy of motion is considered kinetic energy.
b. Because, photons exist on the molecular level while electrons do not exist on the molecular level, all energy of motion is considered kinetic energy.
c. Because, electrons exist on the molecular level while photons do not exist on the molecular level, all energy of motion is considered kinetic energy.
d. Because, photons and electrons exist on the molecular level, all energy of motion is considered kinetic energy.

19. In what region of the electromagnetic spectrum will photons be most effective in accelerating a solar sail?

a. ultraviolet rays
b. infrared rays
c. X-rays
d. gamma rays

20. True or false—Electron microscopes can resolve images that are smaller than the images resolved by light microscopes.

a. false
b. true

21. How would observations of Compton scattering change if ultraviolet light were used in place of X-rays?

a. Ultraviolet light carries less energy than X-rays. As a result, Compton scattering would be easier to detect.
b. Ultraviolet light carries less energy than X-rays. As a result, Compton scattering would be more difficult to detect.
c. Ultraviolet light carries more energy than X-rays. As a result, Compton scattering would be easier to detect.
d. Ultraviolet light has higher energy than X-rays. As a result, Compton scattering would be more difficult to detect.

Problems

21.1 Planck and Quantum Nature of Light

22. How many X-ray photons per second are created by an X-ray tube that produces a flux of X-rays having a power of 1.00 W? Assume the average energy per photon is 75.0 keV.

a. 8.33×10^{15} photons
b. 9.1×10^{7} photons
c. 9.1×10^{8} photons
d. 8.33×10^{13} photons

23. What is the frequency of a photon produced in a CRT using a 25.0-kV accelerating potential? This is similar to the layout as in older color television sets.

a. 6.04×10^{-48} Hz
b. 2.77×10^{-48} Hz
c. 3.02×10^{18} Hz
d. 6.04×10^{18} Hz

21.2 Einstein and the Photoelectric Effect

24. What is the binding energy in eV of electrons in magnesium, if the longest-wavelength photon that can eject electrons is 337 nm?

a. 7.44×10^{-19} J
b. 7.44×10^{-49} J
c. 5.90×10^{-17} J
d. 5.90×10^{-19} J

25. Photoelectrons from a material with a binding energy of 2.71 eV are ejected by 420-nm photons. Once ejected, how long does it take these electrons to travel 2.50 cm to a detection device?

a. 8.5×10^{-6} s
b. 3.5×10^{-7} s
c. 43.5×10^{-9} s
d. 8.5×10^{-8} s

21.3 The Dual Nature of Light

26. What is the momentum of a 0.0100-nm-wavelength photon that could detect details of an atom?

a. 6.626×10^{-27} kg · m/s
b. 6.626×10^{-32} kg · m/s
c. 6.626×10^{-34} kg · m/s
d. 6.626×10^{-23} kg · m/s

27. The momentum of light is exactly reversed when reflected straight back from a mirror, assuming negligible recoil of the mirror. Thus the change in

momentum is twice the initial photon momentum. Suppose light of intensity 1.00 kW/m2 reflectsfrom a mirror of area 2.00 m2 each second. Using the most general form of Newton's second law, what is the force on the mirror?

a. 1.33×10^{-5} N
b. 1.33×10^{-6} N
c. 1.33×10^{-7} N
d. 1.33×10^{-8} N

Performance Task

21.3 The Dual Nature of Light

28. Our scientific understanding of light has changed over time. There is evidence to support the wave model of light, just as there is evidence to support the particle model of light.

1. Construct a demonstration that supports the wave model of light. Note—One possible method is to use a piece of aluminum foil, razor blade, and laser to demonstrate wave interference. Can you arrange these materials to create an effective demonstration? In writing, explain how evidence from your demonstration supports the wave model of light.
2. Construct a demonstration that supports the particle model of light. Note—One possible method is to use a negatively charged electroscope, zinc plate, and three light sources of different frequencies. A red laser, a desk lamp, and ultraviolet lamp are typically used. Can you arrange these materials to demonstrate the photoelectric effect? In writing, explain how evidence from your demonstration supports the particle model of light.

TEST PREP

Multiple Choice

21.1 Planck and Quantum Nature of Light

29. A perfect blackbody is a perfect absorber of energy transferred by what method?

a. conduction
b. convection
c. induction
d. radiation

30. Which of the following is a physical entity that is quantized?

a. electric charge of an ion
b. frequency of a sound
c. speed of a car

31. Find the energy in joules of photons of radio waves that leave an FM station that has a 90.0-MHz broadcast frequency.

a. 1.8×10^{-25} J
b. 1.11×10^{-25} J
c. 7.1×10^{-43} J
d. 5.96×10^{-26} J

32. Which region of the electromagnetic spectrum will provide photons of the least energy?

a. infrared light
b. radio waves
c. ultraviolet light
d. X-rays

33. A hot, black coffee mug is sitting on a kitchen table in a dark room. Because it cannot be seen, one assumes that it is not emitting energy in the form of light. Explain the fallacy in this logic.

a. Not all heat is in the form of light energy.
b. Not all light energy falls in the visible portion of the electromagnetic spectrum.
c. All heat is in the form of light energy.
d. All light energy falls in the visible portion of the electromagnetic spectrum.

34. Given two stars of equivalent size, which will have a greater temperature: a red dwarf or a yellow dwarf? Explain. Note—Our sun is considered a yellow dwarf.

a. a yellow dwarf, because yellow light has lower frequency
b. a red dwarf, because red light has lower frequency
c. a red dwarf, because red light has higher frequency
d. a yellow dwarf, because yellow light has higher frequency

21.2 Einstein and the Photoelectric Effect

35. What is a quantum of light called?

a. electron
b. neutron
c. photon
d. proton

36. Which of the following observations from the photoelectric effect is not a violation of classical physics?

a. Electrons are ejected immediately after impact from light.
b. Light can eject electrons from a semi-conductive

material.
c. Light intensity does not influence the kinetic energy of ejected electrons.
d. No electrons are emitted if the light frequency is too low.

37. If 5 eV of energy is supplied to an electron with a binding energy of 2.3 eV, with what kinetic energy will the electron be launched?
a. 2.3 eV
b. 7.3 eV
c. 11.5 eV
d. 2.7 eV

38. Which of the following terms translates to *light-producing voltage*?
a. photoelectric
b. quantum mechanics
c. photoconductive
d. photovoltaic

39. Why is high frequency EM radiation considered more dangerous than long wavelength EM radiation?
a. Long wavelength EM radiation photons carry less energy and therefore have greater ability to disrupt materials through the photoelectric effect.
b. Long wavelength EM radiation photons carry more energy and therefore have greater ability to disrupt materials through the photoelectric effect.
c. High frequency EM radiation photons carry less energy and therefore have lower ability to disrupt materials through the photoelectric effect.
d. High frequency EM radiation photons carry more energy and therefore have greater ability to disrupt materials through the photoelectric effect.

40. Why are UV, X-rays, and gamma rays considered ionizing radiation?
a. UV, X-rays, and gamma rays are capable of ejecting photons from a surface.
b. UV, X-rays, and gamma rays are capable of ejecting neutrons from a surface.
c. UV, X-rays, and gamma rays are capable of ejecting protons from a surface.
d. UV, X-rays, and gamma rays are capable of ejecting electrons from a surface.

21.3 The Dual Nature of Light

41. What two particles interact in Compton scattering?
a. photon and electron
b. proton and electron
c. neutron and electron
d. proton and neutron

42. What is the momentum of a 500-nm photon?
a. $8.35 \times 10^{-26}\text{ kg} \cdot \text{m/s}$
b. $3.31 \times 10^{-40}\text{ kg} \cdot \text{m/s}$
c. $7.55 \times 10^{26}\text{ kg} \cdot \text{m/s}$
d. $1.33 \times 10^{-27}\text{ kg} \cdot \text{m/s}$

43. The conservation of what fundamental physics principle is behind the technology of solar sails?
a. charge
b. mass
c. momentum
d. angular momentum

44. Terms like frequency, amplitude, and period are tied to what component of wave-particle duality?
a. neither the particle nor the wave model of light
b. both the particle and wave models of light
c. the particle model of light
d. the wave model of light

45. Why was it beneficial for Compton to scatter electrons using X-rays and not another region of light like microwaves?
a. because X-rays are more penetrating than microwaves
b. because X-rays have lower frequency than microwaves
c. because microwaves have shorter wavelengths than X-rays
d. because X-rays have shorter wavelength than microwaves

Short Answer

21.1 Planck and Quantum Nature of Light

46. Scientists once assumed that all frequencies of light were emitted with equal probability. Explain what the blackbody radiation curve would look like if this were the case.
a. The blackbody radiation curve would look like a circular path.
b. The blackbody radiation curve would look like an elliptical path.
c. The blackbody radiation curve would look like a vertical line.
d. The blackbody radiation curve would look like a horizontal line.

47. Because there are more gradations to high frequency radiation than low frequency radiation, scientists also thought it possible that a curve titled the *ultraviolet catastrophe* would occur. Explain what the blackbody radiation curve would look like if this were the case.

a. The curve would steadily increase in intensity with increasing frequency.
b. The curve would steadily decrease in intensity with increasing frequency.
c. The curve would be much steeper than in the blackbody radiation graph.
d. The curve would be much flatter than in the blackbody radiation graph.

48. Energy provided by a light exists in the following quantities: 150 J, 225 J, 300 J. Define one possible quantum of energy and provide an energy state that cannot exist with this quantum.
a. 65 J; 450 J cannot exist
b. 70 J; 450 J cannot exist
c. 75 J; 375 J cannot exist
d. 75 J; 100 J cannot exist

49. Why is Planck's recognition of quantum particles considered the dividing line between classical and modern physics?
a. Planck recognized that energy is quantized, which was in sync with the classical physics concepts but not in agreement with modern physics concepts.
b. Planck recognized that energy is quantized, which was in sync with modern physics concepts but not in agreement with classical physics concepts.
c. Prior to Planck's hypothesis, all the classical physics calculations were valid for subatomic particles, but quantum physics calculations were not valid.
d. Prior to Planck's hypothesis, all the classical physics calculations were not valid for macroscopic particles, but quantum physics calculations were valid.

50. How many 500-mm microwave photons are needed to supply the 8 kJ of energy necessary to heat a cup of water by 10 degrees Celsius?
a. 8.05×10^{28} photons
b. 8.05×10^{26} photons
c. 2.01×10^{26} photons
d. 2.01×10^{28} photons

51. What is the efficiency of a 100-W, 550-nm lightbulb if a photometer finds that 1×10^{20} photons are emitted each second?
a. 101 percent
b. 72 percent
c. 18 percent
d. 36 percent

52. Rank the following regions of the electromagnetic spectrum by the amount of energy provided per photon: gamma, infrared, microwave, ultraviolet, radio, visible, X-ray.
a. radio, microwave, infrared, visible, ultraviolet, X-ray, gamma
b. radio, infrared, microwave, ultraviolet, visible, X-ray, gamma
c. radio, visible, microwave, infrared, ultraviolet, X-ray, gamma
d. radio, microwave, infrared, visible, ultraviolet, gamma, X-ray

53. Why are photons of gamma rays and X-rays able to penetrate objects more successfully than ultraviolet radiation?
a. Photons of gamma rays and X-rays carry with them less energy.
b. Photons of gamma rays and X-rays have longer wavelengths.
c. Photons of gamma rays and X-rays have lower frequencies.
d. Photons of gamma rays and X-rays carry with them more energy.

21.2 Einstein and the Photoelectric Effect

54. According to wave theory, what is necessary to eject electrons from a surface?
a. Enough energy to overcome the binding energy of the electrons at the surface
b. A frequency that is higher than that of the electrons at the surface
c. Energy that is lower than the binding energy of the electrons at the surface
d. A very small number of photons

55. What is the wavelength of EM radiation that ejects 2.00-eV electrons from calcium metal, given that the binding energy is 2.71 eV?
a. 16.1×10^{5} m
b. 6.21×10^{-5} m
c. 9.94×10^{-26} m
d. 2.63×10^{-7} m

56. Find the wavelength of photons that eject 0.100-eV electrons from potassium, given that the binding energy is $2.24\,\text{eV}$.
a. 6.22×10^{-7} m
b. 5.92×10^{-5} m
c. 1.24×10^{-5} m
d. 5.31×10^{-7} m

57. How do solar cells utilize the photoelectric effect?
a. A solar cell converts all photons that it absorbs to electrical energy using the photoelectric effect.
b. A solar cell converts all electrons that it absorbs to electrical energy using the photoelectric effect.
c. A solar cell absorbs the photons with energy less

than the energy gap of the material of the solar cell and converts it to electrical energy using the photoelectric effect.

d. A solar cell absorbs the photons with energy greater than the energy gap of the material of the solar cell and converts it to electrical energy using the photoelectric effect.

58. Explain the advantages of the photoelectric effect to other forms of energy transformation.

a. The photoelectric effect is able to work on the Sun's natural energy.
b. The photoelectric effect is able to work on energy generated by burning fossil fuels.
c. The photoelectric effect can convert heat energy into electrical energy.
d. The photoelectric effect can convert electrical energy into light energy.

21.3 The Dual Nature of Light

59. Upon collision, what happens to the frequency of a photon?

a. The frequency of the photon will drop to zero.
b. The frequency of the photon will remain the same.
c. The frequency of the photon will increase.
d. The frequency of the photon will decrease.

60. How does the momentum of a photon compare to the momentum of an electron of identical energy?

a. Momentum of the photon is greater than the momentum of an electron.
b. Momentum of the photon is less than the momentum of an electron.
c. Momentum of the photon is equal to the momentum of an electron.
d. Momentum of the photon is zero due to zero rest mass but the momentum of an electron is finite.

61. A 500-nm photon strikes an electron and loses 20 percent of its energy. What is the new momentum of the photon?

a. 4.24×10^{-27} kg · m/s
b. 3.18×10^{-27} kg · m/s
c. 2.12×10^{-27} kg · m/s
d. 1.06×10^{-27} kg · m/s

62. A 500-nm photon strikes an electron and loses 20 percent of its energy. What is the speed of the recoiling electron?

a. 7.18×10^{5} m/s
b. 6.18×10^{5} m/s
c. 5.18×10^{5} m/s
d. 4.18×10^{5} m/s

63. When a photon strikes a solar sail, what is the direction of impulse on the photon?

a. parallel to the sail
b. perpendicular to the sail
c. tangential to the sail
d. opposite to the sail

64. What is a fundamental difference between solar sails and sails that are used on sailboats?

a. Solar sails rely on disorganized strikes from light particles, while sailboats rely on disorganized strikes from air particles.
b. Solar sails rely on disorganized strikes from air particles, while sailboats rely on disorganized strikes from light particles.
c. Solar sails rely on organized strikes from air particles, while sailboats rely on organized strikes from light particles.
d. Solar sails rely on organized strikes from light particles, while sailboats rely on organized strikes from air particles.

65. The wavelength of a particle is called the de Broglie wavelength, and it can be found with the equation $p = \frac{h}{\lambda}$.
Yes or no—Can the wavelength of an electron match that of a proton?

a. Yes, a slow-moving electron can achieve the same momentum as a slow-moving proton.
b. No, a fast-moving electron cannot achieve the same momentum, and hence the same wavelength, as a proton.
c. No, an electron can achieve the same momentum, and hence not the same wavelength, as a proton.
d. Yes, a fast-moving electron can achieve the same momentum, and hence have the same wavelength, as a slow-moving proton.

66. Large objects can move with great momentum. Why then is it difficult to see their wave-like nature?

a. Their wavelength is equal to the object's size.
b. Their wavelength is very small compared to the object's size.
c. Their wavelength is very large compared to the object's size.
d. Their frequency is very small compared to the object's size.

Extended Response

21.1 Planck and Quantum Nature of Light

67. Some television tubes are CRTs. They use an approximately 30-kV accelerating potential to send electrons to the screen, where the electrons stimulate phosphors to emit the light that forms the pictures we watch. Would you expect X-rays also to be created? Explain.

a. No, because the full spectrum of EM radiation is not emitted at any temperature.
b. No, because the full spectrum of EM radiation is not emitted at certain temperatures.
c. Yes, because the full spectrum of EM radiation is emitted at any temperature.
d. Yes, because the full spectrum of EM radiation is emitted at certain temperatures.

68. If Planck's constant were large, say 10^{34} times greater than it is, we would observe macroscopic entities to be quantized. Describe the motion of a child's swing under such circumstances.

a. The child would not be able to swing with particular energies.
b. The child could be released from any height.
c. The child would be able to swing with constant velocity.
d. The child could be released only from particular heights.

69. What is the accelerating voltage of an X-ray tube that produces X-rays with the shortest wavelength of 0.0103 nm?

a. 1.21×10^{10} V
b. 2.4×10^{5} V
c. 3.0×10^{-33} V
d. 1.21×10^{5} V

70. Patients in a doctor's office are rightly concerned about receiving a chest X-ray. Yet visible light is also a form of electromagnetic radiation and they show little concern about sitting under the bright lights of the waiting room. Explain this discrepancy.

a. X-ray photons carry considerably more energy so they can harm the patients.
b. X-ray photons carry considerably less energy so they can harm the patients.
c. X-ray photons have considerably longer wavelengths so they cannot harm the patients.
d. X-ray photons have considerably lower frequencies so they can harm the patients.

21.2 Einstein and the Photoelectric Effect

71. When increasing the intensity of light shining on a metallic surface, it is possible to increase the current created on that surface. Classical theorists would argue that this is evidence that intensity causes charge to move with a greater kinetic energy. Argue this logic from the perspective of a modern physicist.

a. The increased intensity increases the number of ejected electrons. The increased current is due to the increase in the number of electrons.
b. The increased intensity decreases the number of ejected electrons. The increased current is due to the decrease in the number of electrons ejected.
c. The increased intensity does not alter the number of electrons ejected. The increased current is due to the increase in the kinetic energy of electrons.
d. The increased intensity alters the number of electrons ejected, but an increase in the current is due to an increase in the kinetic energy of electrons.

72. What impact does the quantum nature of electromagnetic radiation have on the understanding of speed at the particle scale?

a. Speed must also be quantized at the particle scale.
b. Speed will not be quantized at the particle scale.
c. Speed must be zero at the particle scale.
d. Speed will be infinite at the particle scale.

73. A 500 nm photon of light strikes a semi-conductive surface with a binding energy of 2 eV. With what velocity will an electron be emitted from the semi-conductive surface?

a. 8.38×10^{5} m/s
b. 9.33×10^{5} m/s
c. 3×10^{8} m/s
d. 4.11×10^{5} m/s

74. True or false—Treating food with ionizing radiation helps keep it from spoiling.

a. true
b. false

21.3 The Dual Nature of Light

75. When testing atomic bombs, scientists at Los Alamos recognized that huge releases of energy resulted in problems with power and communications systems in the area surrounding the blast site. Explain the possible tie to Compton scattering.

a. The release of light energy caused large-scale emission of electrons.
b. The release of light energy caused large-scale emission of protons.
c. The release of light energy caused large-scale emission of neutrons.
d. The release of light energy caused large-scale

emission of photons.

76. Sunlight above the Earth's atmosphere has an intensity of 1.30 kW/m2 . If this is reflected straight back from a mirror that has only a small recoil, the light's momentum is exactly reversed, giving the mirror twice the incident momentum. If the mirror were attached to a solar sail craft, how fast would the craft be moving after 24 hr? Note—The average mass per square meter of the craft is 0.100 kg.

a. 8.67×10^{-5} m/s^2
b. 8.67×10^{-6} m/s^2
c. 94.2 m/s
d. 7.49 m/s

77. Consider the counter-clockwise motion of LightSail-1 around Earth. When will the satellite move the fastest?

a. point A
b. point B
c. point C
d. point D

78. What will happen to the interference pattern created by electrons when their velocities are increased?

a. There will be more zones of constructive interference and fewer zones of destructive interference.
b. There will be more zones of destructive interference and fewer zones of constructive interference.
c. There will be more zones of constructive and destructive interference.
d. There will be fewer zones of constructive and destructive interference.

CHAPTER 22
The Atom

Figure 22.1 Individual carbon atoms are visible in this image of a carbon nanotube made by a scanning tunneling electron microscope. (credit: Taner Yildirim, National Institute of Standards and Technology, Wikimedia Commons)

Chapter Outline

22.1 The Structure of the Atom

22.2 Nuclear Forces and Radioactivity

22.3 Half Life and Radiometric Dating

22.4 Nuclear Fission and Fusion

22.5 Medical Applications of Radioactivity: Diagnostic Imaging and Radiation

INTRODUCTION From childhood on, we learn that atoms are a substructure of all things around us, from the air we breathe to the autumn leaves that blanket a forest trail. Invisible to the eye, the atoms have properties that are used to explain many phenomena—a theme found throughout this text. In this chapter, we discuss the discovery of atoms and their own substructures. We will then learn about the forces that keep them together and the tremendous energy they release when we break them apart. Finally, we will see how the knowledge and manipulation of atoms allows us to better understand geology, biology, and the world around us.

22.1 The Structure of the Atom

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Describe Rutherford's experiment and his model of the atom
- Describe emission and absorption spectra of atoms
- Describe the Bohr model of the atom
- Calculate the energy of electrons when they change energy levels
- Calculate the frequency and wavelength of emitted photons when electrons change energy levels
- Describe the quantum model of the atom

Section Key Terms

energy-level diagram	excited state	Fraunhofer lines
ground state	Heisenberg Uncertainty Principle	hydrogen-like atoms
planetary model of the atom	Rutherford scattering	Rydberg constant

How do we know that atoms are really there if we cannot see them with our own eyes? While often taken for granted, our knowledge of the existence and structure of atoms is the result of centuries of contemplation and experimentation. The earliest known speculation on the atom dates back to the fifth century B.C., when Greek philosophers Leucippus and Democritus contemplated whether a substance could be divided without limit into ever smaller pieces. Since then, scientists such as John Dalton (1766–1844), Amadeo Avogadro (1776–1856), and Dmitri Mendeleev (1834–1907) helped to discover the properties of that fundamental structure of matter. While much could be written about any number of important scientific philosophers, this section will focus on the role played by Ernest Rutherford (1871–1937). Though his understanding of our most elemental matter is rooted in the success of countless prior investigations, his surprising discovery about the interior of the atom is most fundamental in explaining so many well-known phenomena.

Rutherford's Experiment

In the early 1900's, the *plum pudding* model was the accepted model of the atom. Proposed in 1904 by J. J. Thomson, the model suggested that the atom was a spherical ball of positive charge, with negatively charged electrons scattered evenly throughout. In that model, the positive charges made up the pudding, while the electrons acted as isolated plums. During its short life, the model could be used to explain why most particles were neutral, although with an unbalanced number of plums, electrically charged atoms could exist.

When Ernest Rutherford began his gold foil experiment in 1909, it is unlikely that anyone would have expected that the plum pudding model would be challenged. However, using a radioactive source, a thin sheet of gold foil, and a phosphorescent screen, Rutherford would uncover something so great that he would later call it "the most incredible event that has ever happened to me in my life"[James, L. K. (1993). *Nobel Laureates in Chemistry, 1901–1992*. Washington, DC: American Chemical Society.]

The experiment that Rutherford designed is shown in Figure 22.2. As you can see in, a radioactive source was placed in a lead container with a hole in one side to produce a beam of positively charged helium particles, called *alpha particles*. Then, a thin gold foil sheet was placed in the beam. When the high-energy alpha particles passed through the gold foil, they were scattered. The scattering was observed from the bright spots they produced when they struck the phosphor screen.

Figure 22.2 Rutherford's experiment gave direct evidence for the size and mass of the nucleus by scattering alpha particles from a thin gold foil. The scattering of particles suggests that the gold nuclei are very small and contain nearly all of the gold atom's mass. Particularly significant in showing the size of the nucleus are alpha particles that scatter to very large angles, much like a soccer ball bouncing off a goalie's head.

The expectation of the plum pudding model was that the high-energy alpha particles would be scattered only slightly by the presence of the gold sheet. Because the energy of the alpha particles was much higher than those typically associated with atoms, the alpha particles should have passed through the thin foil much like a supersonic bowling ball would crash through a

few dozen rows of bowling pins. Any deflection was expected to be minor, and due primarily to the electrostatic Coulomb force between the alpha particles and the foil's interior electric charges.

However, the true result was nothing of the sort. While the majority of alpha particles passed through the foil unobstructed, Rutherford and his collaborators Hans Geiger and Ernest Marsden found that alpha particles occasionally were scattered to large angles, and some even came back in the direction from which they came! The result, called **Rutherford scattering**, implied that the gold nuclei were actually very small when compared with the size of the gold atom. As shown in Figure 22.3, the dense nucleus is surrounded by mostly empty space of the atom, an idea verified by the fact that only 1 in 8,000 particles was scattered backward.

Figure 22.3 An expanded view of the atoms in the gold foil in Rutherford's experiment. Circles represent the atoms that are about 10^{-10} m in diameter, while the dots represent the nuclei that are about 10^{-15} m in diameter. To be visible, the dots are much larger than scale—if the nuclei were actually the size of the dots, each atom would have a diameter of about five meters! Most alpha particles crash through but are relatively unaffected because of their high energy and the electron's small mass. Some, however, strike a nucleus and are scattered straight back. A detailed analysis of their interaction gives the size and mass of the nucleus.

Although the results of the experiment were published by his colleagues in 1909, it took Rutherford two years to convince himself of their meaning. Rutherford later wrote: "It was almost as incredible as if you fired a 15-inch shell at a piece of tissue paper and it came back and hit you. On consideration, I realized that this scattering backwards … [meant] … the greatest part of the mass of the atom was concentrated in a tiny nucleus." In 1911, Rutherford published his analysis together with a proposed model of the atom, which was in part based on Geiger's work from the previous year. As a result of the paper, the size of the nucleus was determined to be about 10^{-15} m, or 100,000 times smaller than the atom. That implies a huge density, on the order of 10^{15} g/cm^3, much greater than any macroscopic matter.

Based on the size and mass of the nucleus revealed by his experiment, as well as the mass of electrons, Rutherford proposed the **planetary model of the atom**. The planetary model of the atom pictures low-mass electrons orbiting a large-mass nucleus. The sizes of the electron orbits are large compared with the size of the nucleus, and most of the atom is a vacuum. The model is analogous to how low-mass planets in our solar system orbit the large-mass Sun. In the atom, the attractive Coulomb force is analogous to gravitation in the planetary system (see Figure 22.4).

Figure 22.4 Rutherford's planetary model of the atom incorporates the characteristics of the nucleus, electrons, and the size of the atom. The model was the first to recognize the structure of atoms, in which low-mass electrons orbit a very small, massive nucleus in orbits much larger than the nucleus. The atom is mostly empty and is analogous to our planetary system.

Virtual Physics

Rutherford Scattering

Click to view content (https://www.openstax.org/l/28rutherford)

How did Rutherford figure out the structure of the atom without being able to see it? Explore the answer through this simulation of the famous experiment in which he disproved the plum pudding model by observing alpha particles bouncing off atoms and determining that they must have a small core.

TIPS FOR SUCCESS

As you progress through the model of the atom, consider the effect that experimentation has on the scientific process. Ask yourself the following: What would our model of the atom be without Rutherford's gold foil experiment? What further understanding of the atom would not have been gained? How would that affect our current technologies? Though often confusing, experiments taking place today to further understand composition of the atom could perhaps have a similar effect.

Absorption and Emission Spectra

In 1900, Max Planck recognized that all energy radiated from a source is emitted by atoms in quantum states. How would that radical idea relate to the interior of an atom? The answer was first found by investigating the spectrum of light or emission spectrum produced when a gas is highly energized.

Figure 22.5 shows how to isolate the emission spectrum of one such gas. The gas is placed in the discharge tube at the left, where it is energized to the point at which it begins to radiate energy or emit light. The radiated light is channeled by a thin slit and then passed through a diffraction grating, which will separate the light into its constituent wavelengths. The separated light will then strike the photographic film on the right.

The line spectrum shown in part (b) of Figure 22.5 is the output shown on the film for excited iron. Note that this spectrum is not continuous but discrete. In other words, only particular wavelengths are emitted by the iron source. Why would that be the case?

Figure 22.5 Part (a) shows, from left to right, a discharge tube, slit, and diffraction grating producing a line spectrum. Part (b) shows the emission spectrum for iron. The discrete lines imply quantized energy states for the atoms that produce them. The line spectrum for each element is unique, providing a powerful and much-used analytical tool, and many line spectra were well known for many years before they could be explained with physics. (credit:(b) Yttrium91, Wikimedia Commons)

The spectrum of light created by excited iron shows a variety of discrete wavelengths emitted within the visible spectrum. Each element, when excited to the appropriate degree, will create a discrete emission spectrum as in part (b) of Figure 22.5. However, the wavelengths emitted will vary from element to element. The emission spectrum for iron was chosen for Figure 22.5 solely

because a substantial portion of its emission spectrum is within the visible spectrum. Figure 22.6 shows the emission spectrum for hydrogen. Note that, while discrete, a large portion of hydrogen emission takes place in the ultraviolet and infrared regions.

Figure 22.6 A schematic of the hydrogen spectrum shows several series named for those who contributed most to their determination. Part of the Balmer series is in the visible spectrum, while the Lyman series is entirely in the ultraviolet, and the Paschen series and others are in the infrared. Values of n_f and n_i are shown for some of the lines. Their importance will be described shortly.

Just as an emission spectrum shows all discrete wavelengths emitted by a gas, an absorption spectrum will show all light that is absorbed by a gas. Black lines exist where the wavelengths are absorbed, with the remainder of the spectrum lit by light is free to pass through. What relationship do you think exists between the black lines of a gas's absorption spectrum and the colored lines of its emission spectrum? Figure 22.7 shows the absorption spectrum of the Sun. The black lines are called **Fraunhofer lines**, and they correspond to the wavelengths absorbed by gases in the Sun's exterior.

Figure 22.7 The absorption spectrum of the Sun. The black lines appear at wavelengths absorbed by the Sun's gas exterior. The energetic photons emitted from the Sun's interior are absorbed by gas in its exterior and reemitted in directions away from the observer. That results in dark lines within the absorption spectrum. The lines are called Fraunhofer lines, in honor of the German physicist who discovered them. Lines similar to those are used to determine the chemical composition of stars well outside our solar system.

Bohr's Explanation of the Hydrogen Spectrum

To tie the unique signatures of emission spectra to the composition of the atom itself would require clever thinking. Niels Bohr (1885–1962), a Danish physicist, did just that, by making immediate use of Rutherford's planetary model of the atom. Bohr, shown in Figure 22.8, became convinced of its validity and spent part of 1912 at Rutherford's laboratory. In 1913, after returning to Copenhagen, he began publishing his theory of the simplest atom, hydrogen, based on Rutherford's planetary model.

Figure 22.8 Niels Bohr, Danish physicist, used the planetary model of the atom to explain the atomic spectrum and size of the hydrogen

atom. His many contributions to the development of atomic physics and quantum mechanics, his personal influence on many students and colleagues, and his personal integrity, especially in the face of Nazi oppression, earned him a prominent place in history. (credit: Unknown Author, Wikimedia Commons)

Bohr was able to derive the formula for the hydrogen spectrum using basic physics, the planetary model of the atom, and some very important new conjectures. His first conjecture was that only certain orbits are allowed: In other words, in an atom, the orbits of electrons are quantized. Each quantized orbit has a different distinct energy, and electrons can move to a higher orbit by absorbing energy or drop to a lower orbit by emitting energy. Because of the quantized orbits, the amount of energy emitted or absorbed must also be quantized, producing the discrete spectra seen in Figure 22.5 and Figure 22.7. In equation form, the amount of energy absorbed or emitted can be found as

$$\Delta E = E_i - E_f, \tag{22.1}$$

where E_i refers to the energy of the initial quantized orbit, and E_f refers to the energy of the final orbits. Furthermore, the wavelength emitted can be found using the equation

$$hf = E_i - E_f, \tag{22.2}$$

and relating the wavelength to the frequency found using the equation $v = f\lambda$, where v corresponds to the speed of light.

It makes sense that energy is involved in changing orbits. For example, a burst of energy is required for a satellite to climb to a higher orbit. What is not expected is that atomic orbits should be quantized. Quantization is not observed for satellites or planets, which can have any orbit, given the proper energy (see Figure 22.9).

Figure 22.9 The planetary model of the atom, as modified by Bohr, has the orbits of the electrons quantized. Only certain orbits are allowed, explaining why atomic spectra are discrete or quantized. The energy carried away from an atom by a photon comes from the electron dropping from one allowed orbit to another and is thus quantized. The same is true for atomic absorption of photons.

Figure 22.10 shows an **energy-level diagram**, a convenient way to display energy states. Each of the horizontal lines corresponds to the energy of an electron in a different orbital. Energy is plotted vertically with the lowest or **ground state** at the bottom and with **excited states** above. The vertical arrow downwards shows energy being emitted out of the atom due to an electron dropping from one excited state to another. That would correspond to a line shown on the atom's emission spectrum. The Lyman series shown in Figure 22.6 results from electrons dropping to the ground state, while the Balmer and Paschen series result to electrons dropping to the $n = 2$ and $n = 3$ states, respectively.

Figure 22.10 An energy-level diagram plots energy vertically and is useful in visualizing the energy states of a system and the transitions between them. This diagram is for the hydrogen-atom electrons, showing a transition between two orbits having energies E_4 and E_2 . The energy transition results in a Balmer series line in an emission spectrum.

Energy and Wavelength of Emitted Hydrogen Spectra

The energy associated with a particular orbital of a hydrogen atom can be found using the equation

$$E_n = -\frac{13.6\text{ eV}}{n^2}(n = 1, 2, 3, \ldots), \tag{22.3}$$

where n corresponds to the orbital value from the atom's nucleus. The negative value in the equation is based upon a baseline energy of zero when the electron is infinitely far from the atom. As a result, the negative value shows that energy is necessary to free the electron from its orbital state. The minimum energy to free the electron is also referred to as its *binding energy*. The equation is only valid for atoms with single electrons in their orbital shells (like hydrogen). For ionized atoms similar to hydrogen, the following formula may be used.

$$E_n = \frac{Z^2}{n^2}E_o\ (n = 1, 2, 3, \ldots) \tag{22.4}$$

Please note that E_o corresponds to −13.6 eV, as mentioned earlier. Additionally, Z refers to the atomic number of the element studied. The atomic number is the number of protons in the nucleus—it is different for each element. The above equation is derived from some basic physics principles, namely conservation of energy, conservation of angular momentum, Coulomb's law, and centripetal force. There are three derivations that result in the orbital energy equations, and they are shown below. While you can use the energy equations without understanding the derivations, they will help to remind you of just how valuable those fundamental concepts are.

Derivation 1 (Finding the Radius of an Orbital)

One primary difference between the planetary model of the solar system and the planetary model of the atom is the cause of the circular motion. While gravitation causes the motion of orbiting planets around an interior star, the Coulomb force is responsible for the circular shape of the electron's orbit. The magnitude of the centripetal force is $\frac{m_e v^2}{r_n}$, while the magnitude of the Coulomb force is $\frac{k(Zq_e)(q_e)}{{r_e}^2}$. The assumption here is that the nucleus is more massive than the stationary electron, and the electron orbits about it. That is consistent with the planetary model of the atom. Equating the Coulomb force and the centripetal force,

$$\frac{m_e v^2}{r_n} = \frac{k(Zq_e)(q_e)}{{r_e}^2}, \tag{22.5}$$

which yields

$$r_n = \frac{k\left(Zq_e{}^2\right)}{mv^2}. \tag{22.6}$$

Derivation 2 (Finding the Velocity of the Orbiting Electron)

Bohr was clever enough to find a way to calculate the electron orbital energies in hydrogen. That was an important first step that has been improved upon, but it is well worth repeating here, because it does correctly describe many characteristics of hydrogen. Assuming circular orbits, Bohr proposed that the angular momentum L of an electron in its orbit is also quantized, that is, it has only specific, discrete values. The value for L is given by the formula

$$L = m_e v r_n = n\frac{h}{2\pi}(n = 1, 2, 3, \ldots), \tag{22.7}$$

where L is the angular momentum, m_e is the electron's mass, r_n is the radius of the n th orbit, and h is Planck's constant. Note that angular momentum is $L = I\omega$. For a small object at a radius r, $I = mr^2$, and $\omega = \frac{v}{r}$, so that $L = I\omega = \left(mr^2\right)\left(\frac{v}{r}\right) = mvr$. Quantization says that the value of mvr can only be equal to h / 2, 2h / 2, 3h / 2, etc. At the time, Bohr himself did not know why angular momentum should be quantized, but by using that assumption, he was able to calculate the energies in the hydrogen spectrum, something no one else had done at the time.

Derivation 3 (Finding the Energy of the Orbiting Electron)

To get the electron orbital energies, we start by noting that the electron energy is the sum of its kinetic and potential energy.

$$E_n = KE + PE \tag{22.8}$$

Kinetic energy is the familiar $KE = \frac{1}{2}mv^2$, assuming the electron is not moving at a relativistic speed. Potential energy for the electron is electrical, or $PE = q_e V$, where V is the potential due to the nucleus, which looks like a point charge. The nucleus has a positive charge Zq_e; thus, $V = \frac{kZq_e}{r_n}$, recalling an earlier equation for the potential due to a point charge from the chapter on Electricity and Magnetism. Since the electron's charge is negative, we see that $PE = \frac{-kZq_e{}^2}{r_n}$. Substituting the expressions for KE and PE,

$$E_n = \frac{1}{2}m_e v^2 - \frac{kZq_e{}^2}{r_n}. \tag{22.9}$$

Now we solve for r_n and v using the equation for angular momentum $L = m_e v r_n = n\frac{h}{2\pi}(n = 1, 2, 3, \ldots)$, giving

$$v = n\frac{h}{2\pi m_e r_n}(n = 1, 2, 3, \ldots) \tag{22.10}$$

and

$$r_n = n\frac{h}{2\pi m_e v}(n = 1, 2, 3, \ldots). \tag{22.11}$$

Substituting the expression for r_n and v into the above expressions for energy (KE and PE), and performing algebraic manipulation, yields

$$E_n = -\frac{Z^2}{n^2}E_o(n = 1, 2, 3, \ldots) \tag{22.12}$$

for the orbital energies of **hydrogen-like atoms**. Here, E_o is the ground-state energy (n = 1) for hydrogen (Z = 1) and is given by

$$E_o = \frac{2\pi^2 q_e{}^4 m_e k^2}{h^2} = 13.6 \text{ eV}. \tag{22.13}$$

Thus, for hydrogen,

$$E_n = -\frac{13.6 \text{ eV}}{n^2}(n = 1, 2, 3, \ldots). \tag{22.14}$$

The relationship between orbital energies and orbital states for the hydrogen atom can be seen in Figure 22.11.

Figure 22.11 Energy-level diagram for hydrogen showing the Lyman, Balmer, and Paschen series of transitions. The orbital energies are calculated using the above equation, first derived by Bohr.

WORKED EXAMPLE

A hydrogen atom is struck by a photon. How much energy must be absorbed from the photon to raise the electron of the hydrogen atom from its ground state to its second orbital?

Strategy

The hydrogen atom has an atomic number of $Z = 1$. Raising the electron from the ground state to its second orbital will increase its orbital level from $n = 1$ to $n = 2$. The energy determined will be measured in electron-volts.

Solution

The amount of energy necessary to cause the change in electron state is the difference between the final and initial energies of the electron. The final energy state of the electron can be found using

$$E_n = \frac{Z^2}{n^2} E_o \; (n = 1, 2, 3, \ldots). \tag{22.15}$$

Knowing the n and Z values for the hydrogen atom, and knowing that $E_o = -13.6$ eV, the result is

$$E_f = \frac{1^2}{2^2}(-13.6 \text{ eV}) = -3.4 \text{ eV}. \tag{22.16}$$

The original amount of energy associated with the electron is equivalent to the ground state orbital, or

$$E_O = \frac{1^2}{1^2}(-13.6 \text{ eV}) = -13.6 \text{ eV}. \tag{22.17}$$

The amount of energy necessary to change the orbital state of the electron can be found by determining the electron's change in energy.

$$\Delta E = E_f - E_o = (-3.4\text{ eV}) - (-13.6\text{ eV}) = +10.2\text{ eV} \tag{22.18}$$

Discussion

The energy required to change the orbital state of the electron is positive. That means that for the electron to move to a state with greater energy, energy must be added to the atom. Should the electron drop back down to its original energy state, a change of −10.2 eV would take place, and 10.2 eV of energy would be emitted from the atom. Just as only quantum amounts of energy may be absorbed by the atom, only quantum amounts of energy can be emitted from the atom. That helps to explain many of the quantum light effects that you have learned about previously.

WORKED EXAMPLE

Characteristic X-Ray Energy

Calculate the approximate energy of an X-ray emitted for an $n = 2$ to $n = 1$ transition in a tungsten anode in an X-ray tube.

Strategy

How do we calculate energies in a multiple-electron atom? In the case of characteristic X-rays, the following approximate calculation is reasonable. Characteristic X-rays are produced when an inner-shell vacancy is filled. Inner-shell electrons are nearer the nucleus than others in an atom and thus feel little net effect from the others. That is similar to what happens inside a charged conductor, where its excess charge is distributed over the surface so that it produces no electric field inside. It is reasonable to assume the inner-shell electrons have hydrogen-like energies, as given by

$$E_n = \frac{Z^2}{n^2} E_o \; (n = 1, 2, 3, \ldots) \tag{22.19}$$

For tungsten, $Z = 74$, so that the effective charge is 73.

Solution

The amount of energy given off as an X-ray is found using

$$\Delta E = hf = E_i - E_f, \tag{22.20}$$

where

$$E_f = -\frac{Z^2}{1^2} E_o = -\frac{73^2}{1}(13.6\text{ eV}) = -72.5\text{ keV} \tag{22.21}$$

and

$$E_i = -\frac{Z^2}{2^2} E_o = -\frac{73^2}{4}(13.6\text{ eV}) = -18.1\text{ keV}. \tag{22.22}$$

Thus,

$$\Delta E = E_i - E_f = (-18.1\text{ keV}) - (-72.5\text{ keV}) = 54.4\text{ keV}. \tag{22.23}$$

Discussion

This large photon energy is typical of characteristic X-rays from heavy elements. It is large compared with other atomic emissions because it is produced when an inner-shell vacancy is filled, and inner-shell electrons are tightly bound. Characteristic X-ray energies become progressively larger for heavier elements because their energy increases approximately as Z^2. Significant accelerating voltage is needed to create such inner-shell vacancies, because other shells are filled and you cannot simply bump one electron to a higher filled shell. You must remove it from the atom completely. In the case of tungsten, at least 72.5 kV is needed. Tungsten is a common anode material in X-ray tubes; so much of the energy of the impinging electrons is absorbed, raising its temperature, that a high-melting-point material like tungsten is required.

The wavelength of light emitted by an atom can also be determined through basic derivations. Let us consider the energy of a photon emitted from a hydrogen atom in a downward transition, given by the equation

$$\Delta E = hf = E_i - E_f \tag{22.24}$$

Substituting $E_n = \left(\frac{-13.6\text{ eV}}{n^2}\right)$, we get

$$hf = (13.6\text{ eV})\left(\frac{1}{{n_f}^2} - \frac{1}{{n_i}^2}\right). \tag{22.25}$$

Dividing both sides of the equation by hc gives us an expression for $\frac{1}{\lambda}$,

$$\frac{hf}{hc} = \frac{f}{c} = \frac{1}{\lambda} = \frac{13.6\text{ eV}}{hc}\left(\frac{1}{{n_f}^2} - \frac{1}{{n_i}^2}\right). \tag{22.26}$$

It can be shown that

$$\left(\frac{13.6\text{ eV}}{hc}\right) = \frac{(13.6\text{ eV})(1.602 \times 10^{-19}\text{J/eV})}{(6.602 \times 10^{-34}\text{J}\cdot\text{s})(2.998 \times 10^{8}\text{m/s})} = 1.097 \times 10^{7}\text{m}^{-1} = R, \tag{22.27}$$

where R is the **Rydberg constant**.

Simplified, the formula for determining emitted wavelength can now be written as

$$\frac{1}{\lambda} = R\left(\frac{1}{{n_f}^2} - \frac{1}{{n_i}^2}\right). \tag{22.28}$$

WORKED EXAMPLE

What wavelength of light is emitted by an electron dropping from the third orbital to the ground state of a hydrogen atom?

Strategy

The ground state of a hydrogen atom is considered the first orbital of the atom. As a result, $n_f = 1$ and $n_i = 3$. The Rydberg constant has already been determined and will be constant regardless of atom chosen.

Solution

$$\frac{1}{\lambda} = R\left(\frac{1}{{n_f}^2} - \frac{1}{{n_i}^2}\right) \tag{22.29}$$

For the equation above, calculate wavelength based on the known energy states.

$$\frac{1}{\lambda} = 1.097 \times 10^{7}\left(\frac{1}{1^2} - \frac{1}{3^2}\right) = 9.751 \times 10^{6}\text{m}^{-1} \tag{22.30}$$

Rearranging the equation for wavelength yields

$$\lambda = 1.026 \times 10^{-7}\text{m} = 102.6\text{ nm}. \tag{22.31}$$

Discussion

This wavelength corresponds to light in the ultraviolet spectrum. As a result, we would not be able to see the photon of light emitted when an electron drops from its third to first energy state. However, it is worth noting that by supplying light of wavelength precisely 102.6 nm, we can cause the electron in hydrogen to move from its first to its third orbital state.

Limits of Bohr's Theory and the Quantum Model of the Atom

There are limits to Bohr's theory. It does not account for the interaction of bound electrons, so it cannot be fully applied to multielectron atoms, even one as simple as the two-electron helium atom. Bohr's model is what we call *semiclassical*. The orbits are quantized (nonclassical) but are assumed to be simple circular paths (classical). As quantum mechanics was developed, it became clear that there are no well-defined orbits; rather, there are clouds of probability. Additionally, Bohr's theory did not explain that some spectral lines are doublets or split into two when examined closely. While we shall examine a few of those aspects of quantum mechanics in more detail, it should be kept in mind that Bohr did not fail. Rather, he made very important steps along the path to greater knowledge and laid the foundation for all of atomic physics that has since evolved.

DeBroglie's Waves

Following Bohr's initial work on the hydrogen atom, a decade was to pass before Louis de Broglie proposed that matter has wave properties. The wave-like properties of matter were subsequently confirmed by observations of electron interference when scattered from crystals. Electrons can exist only in locations where they interfere constructively. How does that affect electrons in atomic orbits? When an electron is bound to an atom, its wavelength must fit into a small space, something like a standing wave on a string (see Figure 22.12). Orbits in which an electron can constructively interfere with itself are allowed. All orbits in which constructive interference cannot occur are not able to exist. Thus, only certain orbits are allowed. The wave nature of an electron, according to de Broglie, is why the orbits are quantized!

Figure 22.12 (a) Standing waves on a string have a wavelength related to the length of the string, allowing them to interfere constructively. (b) If we imagine the string formed into a closed circle, we get a rough idea of how electrons in circular orbits can interfere constructively. (c) If the wavelength does not fit into the circumference, the electron interferes destructively; it cannot exist in such an orbit.

For a circular orbit, constructive interference occurs when the electron's wavelength fits neatly into the circumference, so that wave crests always align with crests and wave troughs align with troughs, as shown in Figure 22.12(b). More precisely, when an integral multiple of the electron's wavelength equals the circumference of the orbit, constructive interference is obtained. In equation form, the condition for constructive interference and an allowed electron orbit is

$$n\lambda_n = 2\pi r_n (n = 1, 2, 3, \ldots), \tag{22.32}$$

where λ_n is the electron's wavelength and r_n is the radius of that circular orbit. Figure 22.13 shows the third and fourth orbitals of a hydrogen atom.

Figure 22.13 The third and fourth allowed circular orbits have three and four wavelengths, respectively, in their circumferences.

Heisenberg Uncertainty

How does determining the location of an electron change its trajectory? The answer is fundamentally important—measurement affects the system being observed. It is impossible to measure a physical quantity exactly, and greater precision in measuring one quantity produces less precision in measuring a related quantity. It was Werner Heisenberg who first stated that limit to knowledge in 1929 as a result of his work on quantum mechanics and the wave characteristics of all particles (see Figure 22.14).

Figure 22.14 Werner Heisenberg was the physicist who developed the first version of true quantum mechanics. Not only did his work give a description of nature on the very small scale, it also changed our view of the availability of knowledge. Although he is universally recognized for the importance of his work by receiving the Nobel Prize in 1932, for example, Heisenberg remained in Germany during World War II and headed the German effort to build a nuclear bomb, permanently alienating himself from most of the scientific community. (credit: Unknown Author, Wikimedia Commons)

For example, you can measure the position of a moving electron by scattering light or other electrons from it. However, by doing so, you are giving the electron energy, and therefore imparting momentum to it. As a result, the momentum of the electron is affected and cannot be determined precisely. This change in momentum could be anywhere from close to zero up to the relative momentum of the electron ($p \approx h/\lambda$). Note that, in this case, the particle is an electron, but the principle applies to any particle.

Viewing the electron through the model of wave-particle duality, Heisenberg recognized that, because a wave is not located at one fixed point in space, there is an uncertainty associated with any electron's position. That uncertainty in position, Δx , is approximately equal to the wavelength of the particle. That is, $\Delta x \approx \lambda$. There is an interesting trade-off between position and momentum. The uncertainty in an electron's position can be reduced by using a shorter-wavelength electron, since $\Delta x \approx \lambda$. But shortening the wavelength increases the uncertainty in momentum, since $\Delta p \approx h/\lambda$. Conversely, the uncertainty in momentum can be reduced by using a longer-wavelength electron, but that increases the uncertainty in position. Mathematically, you can express the trade-off by multiplying the uncertainties. The wavelength cancels, leaving

$$\Delta x \Delta p \approx h.$$

Therefore, if one uncertainty is reduced, the other must increase so that their product is $\approx h$. With the use of advanced mathematics, Heisenberg showed that the best that can be done in a simultaneous measurement of position and momentum is

$$\Delta x \Delta p \geq \frac{h}{4\pi}. \tag{22.33}$$

That relationship is known as the **Heisenberg uncertainty principle**.

The Quantum Model of the Atom

Because of the wave characteristic of matter, the idea of well-defined orbits gives way to a model in which there is a cloud of probability, consistent with Heisenberg's uncertainty principle. Figure 22.15 shows how the principle applies to the ground state of hydrogen. If you try to follow the electron in some well-defined orbit using a probe that has a wavelength small enough to

measure position accurately, you will instead knock the electron out of its orbit. Each measurement of the electron's position will find it to be in a definite location somewhere near the nucleus. Repeated measurements reveal a cloud of probability like that in the figure, with each speck the location determined by a single measurement. There is not a well-defined, circular-orbit type of distribution. Nature again proves to be different on a small scale than on a macroscopic scale.

Figure 22.15 The ground state of a hydrogen atom has a probability cloud describing the position of its electron. The probability of finding the electron is proportional to the darkness of the cloud. The electron can be closer or farther than the Bohr radius, but it is very unlikely to be a great distance from the nucleus.

Virtual Physics

Models of the Hydrogen Atom

Click to view content (https://www.openstax.org/l/28atom_model)

How did scientists figure out the structure of atoms without looking at them? Try out different models by shooting light at the atom. Use this simulation to see how the prediction of the model matches the experimental results.

Check Your Understanding

1. Alpha particles are positively charged. What influence did their charge have on the gold foil experiment?
   a. The positively charged alpha particles were attracted by the attractive electrostatic force from the positive nuclei of the gold atoms.
   b. The positively charged alpha particles were scattered by the attractive electrostatic force from the positive nuclei of the gold atoms.
   c. The positively charged alpha particles were scattered by the repulsive electrostatic force from the positive nuclei of the gold atoms.
   d. The positively charged alpha particles were attracted by the repulsive electrostatic force from the positive nuclei of the gold atoms.

22.2 Nuclear Forces and Radioactivity

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Describe the structure and forces present within the nucleus
- Explain the three types of radiation
- Write nuclear equations associated with the various types of radioactive decay

Section Key Terms

alpha decay	atomic number	beta decay	gamma decay	Geiger tube
isotope	mass number	nucleons	radioactive	radioactive decay

radioactivity scintillator strong nuclear force transmutation

There is an ongoing quest to find the substructures of matter. At one time, it was thought that atoms would be the ultimate substructure. However, just when the first direct evidence of atoms was obtained, it became clear that they have a substructure and a tiny nucleus. The nucleus itself has spectacular characteristics. For example, certain nuclei are unstable, and their decay emits radiations with energies millions of times greater than atomic energies. Some of the mysteries of nature, such as why the core of Earth remains molten and how the Sun produces its energy, are explained by nuclear phenomena. The exploration of radioactivity and the nucleus has revealed new fundamental particles, forces, and conservation laws. That exploration has evolved into a search for further underlying structures, such as quarks. In this section, we will explore the fundamentals of the nucleus and nuclear radioactivity.

The Structure of the Nucleus

At this point, you are likely familiar with the neutron and proton, the two fundamental particles that make up the nucleus of an atom. Those two particles, collectively called **nucleons**, make up the small interior portion of the atom. Both particles have nearly the same mass, although the neutron is about two parts in 1,000 more massive. The mass of a proton is equivalent to 1,836 electrons, while the mass of a neutron is equivalent to that of 1,839 electrons. That said, each of the particles is significantly more massive than the electron.

When describing the mass of objects on the scale of nucleons and atoms, it is most reasonable to measure their mass in terms of atoms. The atomic mass unit (u) was originally defined so that a neutral carbon atom would have a mass of exactly 12 u. Given that protons and neutrons are approximately the same mass, that there are six protons and six neutrons in a carbon atom, and that the mass of an electron is minuscule in comparison, measuring this way allows for both protons and neutrons to have masses close to 1 u. Table 22.1 shows the mass of protons, neutrons, and electrons on the new scale.

TIPS FOR SUCCESS

For most conceptual situations, the difference in mass between the proton and neutron is insubstantial. In fact, for calculations that require fewer than four significant digits, both the proton and neutron masses may be considered equivalent to one atomic mass unit. However, when determining the amount of energy released in a nuclear reaction, as in Equation 22.40, the difference in mass cannot be ignored.

Another other useful mass unit on the atomic scale is the MeV/c^2 . While rarely used in most contexts, it is convenient when one uses the equation $E = mc^2$, as will be addressed later in this text.

	Proton Mass	Neutron Mass	Electron Mass
Kilograms (kg)	1.673×10^{-27}	1.675×10^{-27}	9.109×10^{-31}
Atomic mass units (u)	1.007	1.009	5.486×10^{-4}

Table 22.1 Atomic Masses for Multiple Units

To more completely characterize nuclei, let us also consider two other important quantities: the atomic number and the mass number. The **atomic number**, Z, represents the number of protons within a nucleus. That value determines the elemental quality of each atom. Every carbon atom, for instance, has a Z value of 6, whereas every oxygen atom has a Z value of 8. For clarification, only oxygen atoms may have a Z value of 8. If the Z value is not 8, the atom cannot be oxygen.

The **mass number**, A, represents the total number of protons and neutrons, or nucleons, within an atom. For an ordinary carbon atom the mass number would be 12, as there are typically six neutrons accompanying the six protons within the atom. In the case of carbon, the mass would be exactly 12 u. For oxygen, with a mass number of 16, the atomic mass is 15.994915 u. Of course, the difference is minor and can be ignored for most scenarios. Again, because the mass of an electron is so small compared to the nucleons, the mass number and the atomic mass can be essentially equivalent. Figure 22.16 shows an example of Lithium-7, which has an atomic number of 3 and a mass number of 7.

How does the mass number help to differentiate one atom from another? If each atom of carbon has an atomic number of 6,

then what is the value of including the mass number at all? The intent of the mass number is to differentiate between various isotopes of an atom. The term **isotope** refers to the variation of atoms based upon the number of neutrons within their nucleus. While it is most common for there to be six neutrons accompanying the six protons within a carbon atom, it is possible to find carbon atoms with seven neutrons or eight neutrons. Those carbon atoms are respectively referred to as carbon-13 and carbon-14 atoms, with their mass numbers being their primary distinction. The isotope distinction is an important one to make, as the number of neutrons within an atom can affect a number of its properties, not the least of which is nuclear stability.

Figure 22.16 Lithium-7 has three protons and four neutrons within its nucleus. As a result, its mass number is 7, while its atomic number is 3. The actual mass of the atom is 7.016 u. Lithium 7 is an isotope of lithium.

To more easily identify various atoms, their atomic number and mass number are typically written in a form of representation called the nuclide. The nuclide form appears as follows: ${}^{Z}_{A}X_{N}$, where X is the atomic symbol and N represents the number of neutrons.

Let us look at a few examples of nuclides expressed in the ${}^{Z}_{A}X_{N}$ notation. The nucleus of the simplest atom, hydrogen, is a single proton, or ${}^{1}_{1}\mathrm{H}$ (the zero for no neutrons is often omitted). To check the symbol, refer to the periodic table—you see that the atomic number Z of hydrogen is 1. Since you are given that there are no neutrons, the mass number A is also 1. There is a scarce form of hydrogen found in nature called *deuterium*; its nucleus has one proton and one neutron and, hence, twice the mass of common hydrogen. The symbol for deuterium is, thus, ${}^{2}_{1}\mathrm{H}_{2}$. An even rarer—and radioactive—form of hydrogen is called *tritium*, since it has a single proton and two neutrons, and it is written ${}^{3}_{1}\mathrm{H}_{2}$. The three varieties of hydrogen have nearly identical chemistries, but the nuclei differ greatly in mass, stability, and other characteristics. Again, the different nuclei are referred to as isotopes of the same element.

There is some redundancy in the symbols A, X, Z, and N. If the element X is known, then Z can be found in a periodic table. If both A and X are known, then N can also be determined by first finding Z; then, $N = A - Z$. Thus the simpler notation for nuclides is

$$ {}^{A}X, \tag{22.34} $$

which is sufficient and is most commonly used. For example, in this simpler notation, the three isotopes of hydrogen are ${}^{1}\mathrm{H}$, ${}^{2}\mathrm{H}$, and ${}^{3}\mathrm{H}$. For ${}^{238}\mathrm{U}$, should we need to know, we can determine that $Z = 92$ for uranium from the periodic table, and thus, $N = 238 - 92 = 146$.

Radioactivity and Nuclear Forces

In 1896, the French physicist Antoine Henri Becquerel (1852–1908) noticed something strange. When a uranium-rich mineral called pitchblende was placed on a completely opaque envelope containing a photographic plate, it darkened spots on the photographic plate.. Becquerel reasoned that the pitchblende must emit invisible rays capable of penetrating the opaque material. Stranger still was that no light was shining on the pitchblende, which means that the pitchblende was emitting the invisible rays continuously without having any energy input! There is an apparent violation of the law of conservation of energy, one that scientists can now explain using Einstein's famous equation $E = mc^2$. It was soon evident that Becquerel's rays originate in the nuclei of the atoms and have other unique characteristics.

To this point, most reactions you have studied have been chemical reactions, which are reactions involving the electrons

surrounding the atoms. However, two types of experimental evidence implied that Becquerel's rays did not originate with electrons, but instead within the nucleus of an atom.

First, the radiation is found to be only associated with certain elements, such as uranium. Whether uranium was in the form of an element or compound was irrelevant to its radiation. In addition, the presence of radiation does not vary with temperature, pressure, or ionization state of the uranium atom. Since all of those factors affect electrons in an atom, the radiation cannot come from electron transitions, as atomic spectra do.

The huge energy emitted during each event is the second piece of evidence that the radiation cannot be atomic. Nuclear radiation has energies on the order of 10^6 eV per event, which is much greater than typical atomic energies that are a few eV, such as those observed in spectra and chemical reactions, and more than ten times as high as the most energetic X-rays.

But why would reactions within the nucleus take place? And what would cause an apparently stable structure to begin emitting energy? Was there something special about Becquerel's uranium-rich pitchblende? To answer those questions, it is necessary to look into the structure of the nucleus. Though it is perhaps surprising, you will find that many of the same principles that we observe on a macroscopic level still apply to the nucleus.

Nuclear Stability

A variety of experiments indicate that a nucleus behaves something like a tightly packed ball of nucleons, as illustrated in Figure 22.17. Those nucleons have large kinetic energies and, thus, move rapidly in very close contact. Nucleons can be separated by a large force, such as in a collision with another nucleus, but strongly resist being pushed closer together. The most compelling evidence that nucleons are closely packed in a nucleus is that the radius of a nucleus, r, is found to be approximately

$$r = r_o A^{1/3}, \tag{22.35}$$

where $r_o = 1.2$ femtometer (fm) and A is the mass number of the nucleus.

Note that $r^3 \propto A$. Since many nuclei are spherical, and the volume of a sphere is $V = \left(\frac{4}{3}\right)\pi r^3$, we see that $V \propto A$—that is, the volume of a nucleus is proportional to the number of nucleons in it. That is what you expect if you pack nucleons so close that there is no empty space between them.

Figure 22.17 Nucleons are held together by nuclear forces and resist both being pulled apart and pushed inside one another. The volume of the nucleus is the sum of the volumes of the nucleons in it, here shown in different colors to represent protons and neutrons.

So what forces hold a nucleus together? After all, the nucleus is very small and its protons, being positive, should exert tremendous repulsive forces on one another. Considering that, it seems that the nucleus would be forced apart, not together!

The answer is that a previously unknown force holds the nucleus together and makes it into a tightly packed ball of nucleons. This force is known as the **strong nuclear force**. The strong force has such a short range that it quickly fall to zero over a distance of only 10^{-15} meters. However, like glue, it is very strong when the nucleons get close to one another.

The balancing of the electromagnetic force with the nuclear forces is what allows the nucleus to maintain its spherical shape. If, for any reason, the electromagnetic force should overcome the nuclear force, components of the nucleus would be projected outward, creating the very radiation that Becquerel discovered!

Understanding why the nucleus would break apart can be partially explained using Table 22.2. The balance between the strong nuclear force and the electromagnetic force is a tenuous one. Recall that the attractive strong nuclear force exists between any two nucleons and acts over a very short range while the weaker repulsive electromagnetic force only acts between protons, although over a larger range. Considering the interactions, an imperfect balance between neutrons and protons can result in a nuclear reaction, with the result of regaining equilibrium.

	Range of Force	Direction	Nucleon Interaction	Magnitude of Force
Electromagnetic Force	Long range, though decreasing by $1/r^2$	Repulsive	Proton –proton repulsion	Relatively small
Strong Nuclear Force	Very short range, essentially zero at 1 femtometer	Attractive	Attraction between any two nucleons	100 times greater than the electromagnetic force

Table 22.2 Comparing the Electromagnetic and Strong Forces

The radiation discovered by Becquerel was due to the large number of protons present in his uranium-rich pitchblende. In short, the large number of protons caused the electromagnetic force to be greater than the strong nuclear force. To regain stability, the nucleus needed to undergo a nuclear reaction called **alpha (α) decay**.

The Three Types of Radiation

Radioactivity refers to the act of emitting particles or energy from the nucleus. When the uranium nucleus emits energetic nucleons in Becquerel's experiment, the radioactive process causes the nucleus to alter in structure. The alteration is called **radioactive decay**. Any substance that undergoes radioactive decay is said to be **radioactive**. That those terms share a root with the term *radiation* should not be too surprising, as they all relate to the transmission of energy.

Alpha Decay

Alpha decay refers to the type of decay that takes place when too many protons exist in the nucleus. It is the most common type of decay and causes the nucleus to regain equilibrium between its two competing internal forces. During alpha decay, the nucleus ejects two protons and two neutrons, allowing the strong nuclear force to regain balance with the repulsive electromagnetic force. The nuclear equation for an alpha decay process can be shown as follows.

$$^{A}_{Z}X_N \rightarrow {}^{A-4}_{Z-2}Y_N + {}^{4}_{2}\text{He} \tag{22.36}$$

Figure 22.18 A nucleus undergoes alpha decay. The alpha particle can be seen as made up of two neutrons and two protons, which constitute a helium-4 atom.

Three things to note as a result of the above equation:

1. By ejecting an alpha particle, the original nuclide decreases in atomic number. That means that Becquerel's uranium nucleus, upon decaying, is actually transformed into thorium, two atomic numbers lower on the periodic table! The process of changing elemental composition is called **transmutation**.
2. Note that the two protons and two neutrons ejected from the nucleus combine to form a helium nucleus. Shortly after decay, the ejected helium ion typically acquires two electrons to become a stable helium atom.
3. Finally, it is important to see that, despite the elemental change, physical conservation still takes place. The mass number of the new element and the alpha particle together equal the mass number of the original element. Also, the net charge of all particles involved remains the same before and after the transmutation.

Beta Decay

Like alpha decay, **beta (β) decay** also takes place when there is an imbalance between neutrons and protons within the nucleus. For beta decay, however, a neutron is transformed into a proton and electron or vice versa. The transformation allows for the total mass number of the atom to remain the same, although the atomic number will increase by one (or decrease by one). Once again, the transformation of the neutron allows for a rebalancing of the strong nuclear and electromagnetic forces. The nuclear

equation for a beta decay process is shown below.

$$^{A}_{Z}X_N \rightarrow {}^{A}_{Z+1}Y_{N-1} + e + v$$

The symbol v in the equation above stands for a high-energy particle called the neutrino. A nucleus may also emit a positron, and in that case Z decreases and N increases. It is beyond the scope of this section and will be discussed in further detail in the chapter on particles. It is worth noting, however, that the mass number and charge in all beta-decay reactions are conserved.

Figure 22.19 A nucleus undergoes beta decay. The neutron splits into a proton, electron, and neutrino. This particular decay is called β^- decay.

Gamma Decay

Gamma decay is a unique form of radiation that does not involve balancing forces within the nucleus. Gamma decay occurs when a nucleus drops from an excited state to the ground state. Recall that such a change in energy state will release energy from the nucleus in the form of a photon. The energy associated with the photon emitted is so great that its wavelength is shorter than that of an X-ray. Its nuclear equation is as follows.

$$^{A}_{Z}X_N \rightarrow X_N + \gamma \tag{22.37}$$

Figure 22.20 A nucleus undergoes gamma decay. The nucleus drops in energy state, releasing a gamma ray.

WORKED EXAMPLE

Creating a Decay Equation

Write the complete decay equation in $^{A}_{Z}X_N$ notation for beta decay producing ^{137}Ba. Refer to the periodic table for values of Z.

Strategy

Beta decay results in an increase in atomic number. As a result, the original (or parent) nucleus, must have an atomic number of one fewer proton.

Solution

The equation for beta decay is as follows

$$^{A}_{Z}X_N \rightarrow {}_{Z+1}^{A}Y_{N-1} + e + v. \tag{22.38}$$

Considering that barium is the product (or daughter) nucleus and has an atomic number of 56, the original nucleus must be of an atomic number of 55. That corresponds to cesium, or Cs.

$$^{137}_{55}\text{Cs}_N \rightarrow {}^{137}_{56}\text{Ba}_{N-1} + e + v \tag{22.39}$$

The number of neutrons in the parent cesium and daughter barium can be determined by subtracting the atomic number from the mass number (137 – 55 for cesium, 137 – 56 for barium). Substitute those values for the *N* and *N – 1* subscripts in the above equation.

$$^{137}_{55}\text{Cs}_{82} \rightarrow {}^{137}_{56}\text{Ba}_{81} + e + \nu \tag{22.40}$$

Discussion

The terms *parent* and *daughter* nucleus refer to the reactants and products of a nuclear reaction. The terminology is not just used in this example, but in all nuclear reaction examples. The cesium-137 nuclear reaction poses a significant health risk, as its chemistry is similar to that of potassium and sodium, and so it can easily be concentrated in your cells if ingested.

WORKED EXAMPLE

Alpha Decay Energy Found from Nuclear Masses

Find the energy emitted in the α decay of ^{239}Pu.

Strategy

Nuclear reaction energy, such as released in α decay, can be found using the equation $E = mc^2$. We must first find Δm, the difference in mass between the parent nucleus and the products of the decay.

The mass of pertinent particles is as follows

^{239}Pu: 239.052157 u

^{235}U: 235.043924 u

^{4}He: 4.002602 u.

Solution

The decay equation for ^{239}Pu is

$$^{239}\text{Pu} \rightarrow {}^{235}\text{U} + {}^{4}\text{He}. \tag{22.41}$$

Determine the amount of mass lost between the parent and daughter nuclei.

$$\begin{aligned} \Delta m &= m\left({}^{239}\text{P u}\right) - \left(m\left({}^{239}\text{U}\right) + m\left({}^{4}\text{He}\right)\right) \\ \Delta m &= 239.052157 \text{ u} - (235.043924 \text{ u} + 4.002602 \text{ u}) \\ \Delta m &= 0.0005631 \text{ u} \end{aligned} \tag{22.42}$$

Now we can find *E* by entering Δm into the equation.

$$E = (\Delta m)\, c^2 = (0.005631 \text{ u})\, c^2 \tag{22.43}$$

And knowing that $1 \text{ u} = 931.5 \text{ meV}/c^2$, we can find that

$$E = (0.005631)\left(931.5 \text{ MeV}/c^2\right)\left(c^2\right) = 5.25 \text{ MeV}. \tag{22.44}$$

Discussion

The energy released in this α decay is in the MeV range, about 10^6 times as great as typical chemical reaction energies, consistent with previous discussions. Most of the energy becomes kinetic energy of the α particle (or ^{4}He nucleus), which moves away at high speed.

The energy carried away by the recoil of the ^{235}U nucleus is much smaller, in order to conserve momentum. The ^{235}U nucleus can be left in an excited state to later emit photons (γ rays). The decay is spontaneous and releases energy, because the products have less mass than the parent nucleus.

Properties of Radiation

The charges of the three radiated particles differ. Alpha particles, with two protons, carry a net charge of +2. Beta particles, with one electron, carry a net charge of –1. Meanwhile, gamma rays are solely photons, or light, and carry no charge. The difference

in charge plays an important role in how the three radiations affect surrounding substances.

Alpha particles, being highly charged, will quickly interact with ions in the air and electrons within metals. As a result, they have a short range and short penetrating distance in most materials. Beta particles, being slightly less charged, have a larger range and larger penetrating distance. Gamma rays, on the other hand, have little electric interaction with particles and travel much farther. Two diagrams below show the importance of difference in penetration. Table 22.3 shows the distance of radiation penetration, and Figure 22.21 shows the influence various factors have on radiation penetration distance.

Type of Radiation	Range
α particles	A sheet of paper, a few cm of air, fractions of a millimeter of tissue
β particles	A thin aluminum plate, tens of cm of tissue
γ rays	Several cm of lead, meters of concrete

Table 22.3 Comparing Ranges of Radioactive Decay

Figure 22.21 The penetration or range of radiation depends on its energy, the material it encounters, and the type of radiation. (a) Greater energy means greater range. (b) Radiation has a smaller range in materials with high electron density. (c) Alphas have the smallest range, betas have a greater range, and gammas have the greatest range.

LINKS TO PHYSICS

Radiation Detectors

The first direct detection of radiation was Becquerel's darkened photographic plate. Photographic film is still the most common detector of ionizing radiation, being used routinely in medical and dental X-rays. Nuclear radiation can also be captured on film, as seen in Figure 22.22. The mechanism for film exposure by radiation is similar to that by photons. A quantum of energy from a radioactive particle interacts with the emulsion and alters it chemically, thus exposing the film. Provided the radiation has more than the few eV of energy needed to induce the chemical change, the chemical alteration will occur. The amount of film darkening is related to the type of radiation and amount of exposure. The process is not 100 percent efficient, since not all incident radiation interacts and not all interactions produce the chemical change.

Figure 22.22 Film badges contain film similar to that used in this dental X-ray film. It is sandwiched between various absorbers to determine the penetrating ability of the radiation as well as the amount. Film badges are worn to determine radiation exposure. (credit: Werneuchen, Wikimedia Commons)

Another very common radiation detector is the **Geiger tube**. The clicking and buzzing sound we hear in dramatizations and documentaries, as well as in our own physics labs, is usually an audio output of events detected by a Geiger counter. These relatively inexpensive radiation detectors are based on the simple and sturdy Geiger tube, shown schematically in Figure 22.23. A conducting cylinder with a wire along its axis is filled with an insulating gas so that a voltage applied between the cylinder and wire produces almost no current. Ionizing radiation passing through the tube produces free ion pairs that are attracted to the wire and cylinder, forming a current that is detected as a count. Not every particle is detected, since some radiation can pass through without producing enough ionization. However, Geiger counters are very useful in producing a prompt output that reveals the existence and relative intensity of ionizing radiation.

(a)

(b)

Figure 22.23 (a) Geiger counters such as this one are used for prompt monitoring of radiation levels, generally giving only relative intensity and not identifying the type or energy of the radiation. (credit: Tim Vickers, Wikimedia Commons) (b) Voltage applied between the cylinder and wire in a Geiger tube affects ions and electrons produced by radiation passing through the gas-filled cylinder. Ions move toward the cylinder and electrons toward the wire. The resulting current is detected and registered as a count.

Another radiation detection method records light produced when radiation interacts with materials. The energy of the radiation is sufficient to excite atoms in a material that may fluoresce, such as the phosphor used by Rutherford's group. Materials called **scintillators** use a more complex process to convert radiation energy into light. Scintillators may be liquid or solid, and they can

be very efficient. Their light output can provide information about the energy, charge, and type of radiation. Scintillator light flashes are very brief in duration, allowing the detection of a huge number of particles in short periods of time. Scintillation detectors are used in a variety of research and diagnostic applications. Among those are the detection of the radiation from distant galaxies using satellite-mounted equipment and the detection of exotic particles in accelerator laboratories.

Virtual Physics

Beta Decay

Click to view content (https://www.openstax.org/l/21betadecayvid)

Watch beta decay occur for a collection of nuclei or for an individual nucleus. With this applet, individuals or groups of students can compare half-lives!

Check Your Understanding

2. What leads scientists to infer that the nuclear strong force exists?
   a. A strong force must hold all the electrons outside the nucleus of an atom.
   b. A strong force must counteract the highly attractive Coulomb force in the nucleus.
   c. A strong force must hold all the neutrons together inside the nucleus.
   d. A strong force must counteract the highly repulsive Coulomb force between protons in the nucleus.

22.3 Half Life and Radiometric Dating

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Explain radioactive half-life and its role in radiometric dating
- Calculate radioactive half-life and solve problems associated with radiometric dating

Section Key Terms

activity	becquerel	carbon-14 dating
decay constant	half-life	radioactive dating

Half-Life and the Rate of Radioactive Decay

Unstable nuclei decay. However, some nuclides decay faster than others. For example, radium and polonium, discovered by Marie and Pierre Curie, decay faster than uranium. That means they have shorter lifetimes, producing a greater rate of decay. Here we will explore half-life and activity, the quantitative terms for lifetime and rate of decay.

Why do we use the term like *half-life* rather than *lifetime*? The answer can be found by examining Figure 22.24, which shows how the number of radioactive nuclei in a sample decreases with time. The time in which half of the original number of nuclei decay is defined as the **half-life**, $t_{1/2}$. After one half-life passes, half of the remaining nuclei will decay in the next half-life. Then, half of that amount in turn decays in the following half-life. Therefore, the number of radioactive nuclei decreases from N to $N/2$ in one half-life, to $N/4$ in the next, to $N/8$ in the next, and so on. Nuclear decay is an example of a purely statistical process.

TIPS FOR SUCCESS

A more precise definition of half-life is that each nucleus has a 50 percent chance of surviving for a time equal to one half-life. If an individual nucleus survives through that time, it still has a 50 percent chance of surviving through another half-life. Even if it happens to survive hundreds of half-lives, it still has a 50 percent chance of surviving through one more. Therefore, the decay of a nucleus is like random coin flipping. The chance of heads is 50 percent, no matter what has happened before. The probability concept aligns with the traditional definition of half-life. Provided the number of nuclei is reasonably large, half of the original nuclei should decay during one half-life period.

Figure 22.24 Radioactive decay reduces the number of radioactive nuclei over time. In one half-life ($t_{1/2}$), the number decreases to half of its original value. Half of what remains decays in the next half-life, and half of that in the next, and so on. This is exponential decay, as seen in the graph of the number of nuclei present as a function of time.

The following equation gives the quantitative relationship between the original number of nuclei present at time zero (N_O) and the number (N) at a later time t

$$N = N_O e^{-\lambda t}, \tag{22.45}$$

where $e = 2.71828\ldots$ is the base of the natural logarithm, and λ is the **decay constant** for the nuclide. The shorter the half-life, the larger is the value of λ, and the faster the exponential $e^{-\lambda t}$ decreases with time. The decay constant can be found with the equation

$$\lambda = \frac{\ln(2)}{t_{1/2}} \approx \frac{0.693}{t_{1/2}}. \tag{22.46}$$

Activity, the Rate of Decay

What do we mean when we say a source is highly radioactive? Generally, it means the number of decays per unit time is very high. We define **activity** R to be the rate of decay expressed in decays per unit time. In equation form, this is

$$R = \frac{\Delta N}{\Delta t}, \tag{22.47}$$

where ΔN is the number of decays that occur in time Δt .

Activity can also be determined through the equation

$$R = \lambda N, \tag{22.48}$$

which shows that as the amount of radiative material (N) decreases, the rate of decay decreases as well.

The SI unit for activity is one decay per second and it is given the name **becquerel** (Bq) in honor of the discoverer of radioactivity. That is,

$$1 \text{ Bq} = 1 \text{ decay/second}.$$

Activity R is often expressed in other units, such as decays per minute or decays per year. One of the most common units for activity is the curie (Ci), defined to be the activity of 1 g of ^{226}Ra, in honor of Marie Curie's work with radium. The definition of the curie is

$$1 \text{ Ci} = 3.70 \times 10^{10} \text{ Bq}, \tag{22.49}$$

or 3.70×10^{10} decays per second.

Radiometric Dating

Radioactive dating or radiometric dating is a clever use of naturally occurring radioactivity. Its most familiar application is **carbon-14 dating**. Carbon-14 is an isotope of carbon that is produced when solar neutrinos strike ^{14}N particles within the atmosphere. Radioactive carbon has the same chemistry as stable carbon, and so it mixes into the biosphere, where it is consumed and becomes part of every living organism. Carbon-14 has an abundance of 1.3 parts per trillion of normal carbon, so if you know the number of carbon nuclei in an object (perhaps determined by mass and Avogadro's number), you can multiply that number by 1.3×10^{-12} to find the number of ^{14}C nuclei within the object. Over time, carbon-14 will naturally decay back to ^{14}N with a half-life of 5,730 years (note that this is an example of beta decay). When an organism dies, carbon exchange with the environment ceases, and ^{14}C is not replenished. By comparing the abundance of ^{14}C in an artifact, such as mummy wrappings, with the normal abundance in living tissue, it is possible to determine the artifact's age (or time since death). Carbon-14 dating can be used for biological tissues as old as 50 or 60 thousand years, but is most accurate for younger samples, since the abundance of ^{14}C nuclei in them is greater.

One of the most famous cases of carbon-14 dating involves the Shroud of Turin, a long piece of fabric purported to be the burial shroud of Jesus (see Figure 22.25). This relic was first displayed in Turin in 1354 and was denounced as a fraud at that time by a French bishop. Its remarkable negative imprint of an apparently crucified body resembles the then-accepted image of Jesus. As a result, the relic has been remained controversial throughout the centuries. Carbon-14 dating was not performed on the shroud until 1988, when the process had been refined to the point where only a small amount of material needed to be destroyed. Samples were tested at three independent laboratories, each being given four pieces of cloth, with only one unidentified piece from the shroud, to avoid prejudice. All three laboratories found samples of the shroud contain 92 percent of the ^{14}C found in living tissues, allowing the shroud to be dated (see Equation 22.57).

Figure 22.25 Part of the Shroud of Turin, which shows a remarkable negative imprint likeness of Jesus complete with evidence of crucifixion wounds. The shroud first surfaced in the 14th century and was only recently carbon-14 dated. It has not been determined how the image was placed on the material. (credit: Butko, Wikimedia Commons)

WORKED EXAMPLE

Carbon-11 Decay

Carbon-11 has a half-life of 20.334 min. (a) What is the decay constant for carbon-11?
If 1 kg of carbon-11 sample exists at the beginning of an hour, (b) how much material will remain at the end of the hour and (c) what will be the decay activity at that time?

Strategy

Since N_O refers to the amount of carbon-11 at the start, then after one half-life, the amount of carbon-11 remaining will be $N_O/2$. The decay constant is equivalent to the probability that a nucleus will decay each second. As a result, the half-life will need to be converted to seconds.

Solution

(a)

$$N = N_O e^{-\lambda t} \tag{22.50}$$

Since half of the carbon-11 remains after one half-life, $N/N_O = 0.5$.

$$0.5 = e^{-\lambda t} \tag{22.51}$$

Take the natural logarithm of each side to isolate the decay constant.

$$\ln(0.5) = -\lambda t \tag{22.52}$$

Convert the 20.334 min to seconds.

$$\begin{aligned} -0.693 &= (-\lambda)(20.334\ \text{min})(\tfrac{60\ \text{s}}{1\ \text{min}}) \\ -0.693 &= (-\lambda)(1{,}220.04\ \text{s}) \\ \tfrac{-0.693}{1{,}220.04\ \text{s}} &= -\lambda \\ \lambda &= 5.68 \times 10^{-4}\text{s}^{-1} \end{aligned} \tag{22.53}$$

(b) The amount of material after one hour can be found by using the equation

$$N = N_O e^{-\lambda t}, \tag{22.54}$$

with t converted into seconds and N_O written as 1,000 g

$$\begin{aligned} N &= (1{,}000\ \text{g})e - (0.000568)(60.60) \\ N &= 129.4\ g \end{aligned} \tag{22.55}$$

(c) The decay activity after one hour can be found by using the equation

$$R = \lambda N \tag{22.56}$$

for the mass value after one hour.

$$R = \lambda N = \left(0.000568\ \frac{\text{decays}}{\text{second}}\right)(129.4\ \text{grams}) = 0.0735\ \text{Bq} \tag{22.57}$$

Discussion

(a) The decay constant shows that 0.0568 percent of the nuclei in a carbon-11 sample will decay each second. Another way of considering the decay constant is that a given carbon-11 nuclei has a 0.0568 percent probability of decaying each second. The decay of carbon-11 allows it to be used in positron emission topography (PET) scans; however, its 20.334 min half-life does pose challenges for its administration.

(b) One hour is nearly three full half-lives of the carbon-11 nucleus. As a result, one would expect the amount of sample remaining to be approximately one eighth of the original amount. The 129.4 g remaining is just a bit larger than one-eighth, which is sensible given a half-life of just over 20 min.

(c) Label analysis shows that the unit of Becquerel is sensible, as there are 0.0735 g of carbon-11 decaying each second. That is smaller amount than at the beginning of the hour, when $R = \left(0.000568\ \frac{\text{decay}}{\text{s}}\right)(1{,}000\ \text{g}) = 0.568$ g of carbon-11 were decaying each second.

WORKED EXAMPLE

How Old is the Shroud of Turin?

Calculate the age of the Shroud of Turin given that the amount of ^{14}C found in it is 92 percent of that in living tissue.

Strategy

Because 92 percent of the ^{14}C remains, $N/N_O = 0.92$. Therefore, the equation $N = N_O e^{-\lambda t}$ can be used to find λt . We also know that the half-life of ^{14}C is 5,730 years, and so once λt is known, we can find λ and then find t as requested. Here, we assume that the decrease in ^{14}C is solely due to nuclear decay.

Solution

Solving the equation $N = N_O e^{-\lambda t}$ for N/N_O gives

$$\frac{N}{N_O} = e^{-\lambda t}. \tag{22.58}$$

Thus,

$$0.92 = e^{-\lambda t}. \tag{22.59}$$

Taking the natural logarithm of both sides of the equation yields

$$\ln\ 0.92 = -\lambda t \tag{22.60}$$

so that

$$-0.0834 = -\lambda t. \tag{22.61}$$

Rearranging to isolate t gives

$$t = \frac{0.0834}{\lambda}. \tag{22.62}$$

Now, the equation $\lambda = \frac{0.693}{t_{1/2}}$ can be used to find λ for ^{14}C. Solving for λ and substituting the known half-life gives

$$\lambda = \frac{0.693}{t_{1/2}} = \frac{0.693}{5{,}730\text{ years}} = 1.21\ \times\ 10^{-4}\text{y}^{-1}. \tag{22.63}$$

We enter that value into the previous equation to find t.

$$t = \frac{0.0834}{1.21\ \times\ 10^{-4}} = 690\text{ years}. \tag{22.64}$$

Discussion

This dates the material in the shroud to 1988 − 690 = 1300. Our calculation is only accurate to two digits, so that the year is rounded to 1300. The values obtained at the three independent laboratories gave a weighted average date of 1320 ± 60. That uncertainty is typical of carbon-14 dating and is due to the small amount of 14 C in living tissues, the amount of material available, and experimental uncertainties (reduced by having three independent measurements). That said, is it notable that the carbon-14 date is consistent with the first record of the shroud's existence and certainly inconsistent with the period in which Jesus lived.

There are other noncarbon forms of radioactive dating. Rocks, for example, can sometimes be dated based on the decay of ^{238}U. The decay series for ^{238}U ends with ^{206}Pb, so the ratio of those nuclides in a rock can be used an indication of how long it has been since the rock solidified. Knowledge of the ^{238}U half-life has shown, for example, that the oldest rocks on Earth solidified about $3.5\ \times\ 10^{9}$ years ago.

Virtual Physics

Radioactive Dating Game

Click to view content (https://www.openstax.org/l/02radioactive_dating_game)

Learn about different types of radiometric dating, such as carbon dating. Understand how decay and half-life work to enable radiometric dating to work. Play a game that tests your ability to match the percentage of the dating element that remains to the age of the object.

22.4 Nuclear Fission and Fusion

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Explain nuclear fission
- Explain nuclear fusion
- Describe how the processes of fission and fusion work in nuclear weapons and in generating nuclear power

Section Key Terms

chain reaction	critical mass	liquid drop model
nuclear fission	nuclear fusion	proton-proton cycle

The previous section dealt with naturally occurring nuclear decay. Without human intervention, some nuclei will change composition in order to achieve a stable equilibrium. This section delves into a less-natural process. Knowing that energy can be emitted in various forms of nuclear change, is it possible to create a nuclear reaction through our own intervention? The answer to this question is yes. Through two distinct methods, humankind has discovered multiple ways of manipulating the atom to release its internal energy.

Nuclear Fission

In simplest terms, **nuclear fission** is the splitting of an atomic bond. Given that it requires great energy separate two nucleons, it may come as a surprise to learn that splitting a nucleus can *release* vast potential energy. And although it is true that huge amounts of energy can be released, considerable effort is needed to do so in practice.

An unstable atom will naturally decay, but it may take millions of years to do so. As a result, a physical catalyst is necessary to produce useful energy through nuclear fission. The catalyst typically occurs in the form of a free neutron, projected directly at the nucleus of a high-mass atom.

As shown in Figure 22.26, a neutron strike can cause the nucleus to elongate, much like a drop of liquid water. This is why the model is known as the **liquid drop model**. As the nucleus elongates, nucleons are no longer so tightly packed, and the repulsive electromagnetic force can overcome the short-range strong nuclear force. The imbalance of forces can result in the two ends of the drop flying apart, with some of the nuclear binding energy released to the surroundings.

Figure 22.26 Neutron-induced fission is shown. First, energy is put into a large nucleus when it absorbs a neutron. Acting like a struck liquid drop, the nucleus deforms and begins to narrow in the middle. Since fewer nucleons are in contact, the repulsive Coulomb force is able to break the nucleus into two parts with some neutrons also flying away.

As you can imagine, the consequences of the nuclei splitting are substantial. When a nucleus is split, it is not only energy that is released, but a small number of neutrons as well. Those neutrons have the potential to cause further fission in other nuclei, especially if they are directed back toward the other nuclei by a dense shield or neutron reflector (see part (d) of Figure 22.26).

However, not every neutron produced by fission induces further fission. Some neutrons escape the fissionable material, while others interact with a nucleus without making it split. We can enhance the number of fissions produced by neutrons by having a large amount of fissionable material as well as a neutron reflector. The minimum amount necessary for self-sustained fission of a given nuclide is called its **critical mass**. Some nuclides, such as ^{239}Pu, produce more neutrons per fission than others, such as ^{235}U. Additionally, some nuclides are easier to make fission than others. In particular, ^{235}U and ^{239}Pu are easier to fission than the much more abundant ^{238}U. Both factors affect critical mass, which is smallest for ^{239}Pu. The self-sustained fission of nuclei is commonly referred to as a **chain reaction**, as shown in Figure 22.27.

Figure 22.27 A chain reaction can produce self-sustained fission if each fission produces enough neutrons to induce at least one more fission. This depends on several factors, including how many neutrons are produced in an average fission and how easy it is to make a particular type of nuclide fission.

A chain reaction can have runaway results. If each atomic split results in two nuclei producing a new fission, the number of nuclear reactions will increase exponentially. One fission will produce two atoms, the next round of fission will create four atoms, the third round eight atoms, and so on. Of course, each time fission occurs, more energy will be emitted, further increasing the power of the atomic reaction. And that is just if two neutrons create fission reactions each round. Perhaps you can now see why so many people consider atomic energy to be an exciting energy source!

To make a self-sustained nuclear fission reactor with ^{235}U, it is necessary to slow down the neutrons. Water is very effective at this, since neutrons collide with protons in water molecules and lose energy. Figure 22.28 shows a schematic of a reactor design called the *pressurized water reactor*.

Figure 22.28 A pressurized water reactor is cleverly designed to control the fission of large amounts of ^{235}U, while using the heat produced in the fission reaction to create steam for generating electrical energy. Control rods adjust neutron flux so that it is self-sustaining. In case the reactor overheats and boils the water away, the chain reaction terminates, because water is needed to slow down the neutrons. This inherent safety feature can be overwhelmed in extreme circumstances.

Control rods containing nuclides that very strongly absorb neutrons are used to adjust neutron flux. To produce large amounts of power, reactors contain hundreds to thousands of critical masses, and the chain reaction easily becomes self-sustaining. Neutron flux must be carefully regulated to avoid an out-of-control exponential increase in the rate of fission.

Control rods help prevent overheating, perhaps even a meltdown or explosive disassembly. The water that is used to slow down neutrons, necessary to get them to induce fission in ^{235}U, and achieve criticality, provides a negative feedback for temperature increase. In case the reactor overheats and boils the water to steam or is breached, the absence of water kills the chain reaction. Considerable heat, however, can still be generated by the reactor's radioactive fission products. Other safety features, thus, need to be incorporated in the event of a loss of coolant accident, including auxiliary cooling water and pumps.

Energies in Nuclear Fission

The following are two interesting facts to consider:

- The average fission reaction produces 200 MeV of energy.
- If you were to measure the mass of the products of a nuclear reaction, you would find that their mass was slightly less than the mass of the original nucleus.

How are those things possible? Doesn't the fission reaction's production of energy violate the conservation of energy? Furthermore, doesn't the loss in mass in the reaction violate the conservation of mass? Those are important questions, and they can both be answered with one of the most famous equations in scientific history.

$$E = mc^2 \tag{22.65}$$

Recall that, according to Einstein's theory, energy and mass are essentially the same thing. In the case of fission, the mass of the products is less than that of the reactants because the missing mass appears in the form of the energy released in the reaction, with a constant value of c^2 Joules of energy converted for each kilogram of material. The value of c^2 is substantial—from Einstein's equation, the amount of energy in just 1 gram of mass would be enough to support the average U.S. citizen for more than 270 years! The example below will show you how a mass-energy transformation of this type takes place.

WORKED EXAMPLE

Calculating Energy from a Kilogram of Fissionable Fuel

Calculate the amount of energy produced by the fission of 1.00 kg of ^{235}U, given the average fission reaction of

$$^{235}\text{U produces 200 MeV.} \tag{22.66}$$

Strategy

The total energy produced is the number of ^{235}U atoms times the given energy per ^{235}U fission. We should therefore find the number of ^{235}U atoms in 1.00 kg.

Solution

The number of ^{235}U atoms in 1.00 kg is Avogadro's number times the number of moles. One mole of ^{235}U has a mass of 235.04 g; thus, there are $(1{,}000\text{ g})/(235.04.00\text{ g/mol}) = 4.25\text{ mol}$. The number of ^{235}U atoms is therefore

$$(4.25\text{ mol})\left(6.02\times10^{23}\text{U/mol}\right) = 2.56\times10^{24}\text{atomsof235U.}$$

So the total energy released is

$$E = \left(2.56\times10^{24}\right)\left(\frac{200\text{ MeV}}{^{235}\text{U}}\right)\left(\frac{1.60\times10^{-13}\text{J}}{\text{MeV}}\right) = 8.21\times10^{13}\text{J.} \tag{22.67}$$

Discussion

The result is another impressively large amount of energy, equivalent to about 14,000 barrels of crude oil or 600,000 gallons of gasoline. But, it is only one fourth the energy produced by the fusion of a kilogram of a mixture of deuterium and tritium. Even though each fission reaction yields about ten times the energy of a fusion reaction, the energy per kilogram of fission fuel is less, because there are far fewer moles per kilogram of the heavy nuclides. Fission fuel is also much scarcer than fusion fuel, and less than 1 percent of uranium (the 235 U) is readily usable.

Virtual Physics

Nuclear Fission

Click to view content (https://www.openstax.org/l/16fission)

Start a chain reaction, or introduce nonradioactive isotopes to prevent one. Use the applet to control energy production in a nuclear reactor!

Nuclear Fusion

Nuclear fusion is defined as the combining, or fusing, of two nuclei and, the combining of nuclei also results in an emission of energy. For many, the concept is counterintuitive. After all, if energy is released when a nucleus is split, how can it also be released when nucleons are combined together? The difference between fission and fusion, which results from the size of the nuclei involved, will be addressed next.

Remember that the structure of a nucleus is based on the interplay of the compressive nuclear strong force and the repulsive electromagnetic force. For nuclei that are less massive than iron, the nuclear force is actually stronger than that of the Coulomb force. As a result, when a low-mass nucleus absorbs nucleons, the added neutrons and protons bind the nucleus more tightly. The increased nuclear strong force does work on the nucleus, and energy is released.

Once the size of the created nucleus exceeds that of iron, the short-ranging nuclear force does not have the ability to bind a nucleus more tightly, and the emission of energy ceases. In fact, for fusion to occur for elements of greater mass than iron, energy must be added to the system! Figure 22.29 shows an energy-mass curve commonly used to describe nuclear reactions. Notice the location of iron (Fe) on the graph. All low-mass nuclei to the left of iron release energy through fusion, while all high-mass particles to the right of iron produce energy through fission.

Figure 22.29 Fusion of light nuclei to form medium-mass nuclei converts mass to energy, because binding energy per nucleon (BE/A) is greater for the product nuclei. The larger BE/A is, the less mass per nucleon, and so mass is converted to energy and released in such fusion reactions.

TIPS FOR SUCCESS

Just as it is not possible for the elements to the left of iron in the figure to naturally fission, it is not possible for elements to the right of iron to naturally undergo fusion, as that process would require the addition of energy to occur. Furthermore, notice that elements commonly discussed in fission and fusion are elements that can provide the greatest change in binding energy, such as uranium and hydrogen.

Iron's location on the energy-mass curve is important, and explains a number of its characteristics, including its role as an elemental endpoint in fusion reactions in stars.

The major obstruction to fusion is the Coulomb repulsion force between nuclei. Since the attractive nuclear force that can fuse nuclei together is short ranged, the repulsion of like positive charges must be overcome in order to get nuclei close enough to induce fusion. Figure 22.30 shows an approximate graph of the potential energy between two nuclei as a function of the distance between their centers. The graph resembles a hill with a well in its center. A ball rolled to the left must have enough kinetic energy to get over the hump before it falls into the deeper well with a net gain in energy. So it is with fusion. If the nuclei are given enough kinetic energy to overcome the electric potential energy due to repulsion, then they can combine, release energy, and fall into a deep well. One way to accomplish that end is to heat fusion fuel to high temperatures so that the kinetic energy of thermal motion is sufficient to get the nuclei together.

Figure 22.30 Potential energy between two light nuclei graphed as a function of distance between them. If the nuclei have enough kinetic energy to get over the Coulomb repulsion hump, they combine, release energy, and drop into a deep attractive well.

You might think that, in our Sun, nuclei are constantly coming into contact and fusing. However, this is only partially true. Only at the Sun's core are the particles close enough and the temperature high enough for fusion to occur!

In the series of reactions below, the Sun produces energy by fusing protons, or hydrogen nuclei ($^1\mathrm{H}$, by far the Sun's most

abundant nuclide) into helium nuclei ^4He. The principal sequence of fusion reactions forms what is called the **proton-proton cycle**

$$\begin{aligned} &^1\text{H} + {}^1\text{H} \rightarrow {}^2\text{H} + e^+ + \nu_e && (0.42\ \text{MeV}) \\ &^1\text{H} + {}^2\text{H} \rightarrow {}^3\text{He} + \gamma && (5.49\ \text{MeV}) \\ &^3\text{He} + {}^3\text{He} \rightarrow {}^4\text{He} + {}^1\text{H} + {}^1\text{H} && (12.86\ \text{MeV}), \end{aligned} \tag{22.68}$$

where e^+ stands for a positron and ν_e is an electron neutrino. The energy in parentheses is *released* by the reaction. Note that the first two reactions must occur twice for the third to be possible, so the cycle consumes six protons (^1H) but gives back two. Furthermore, the two positrons produced will find two electrons and annihilate to form four more γ rays, for a total of six. The overall cycle is thus

$$2e - + 4\,^1\text{H} \rightarrow 4\ \text{He} + 2\nu_e + 6\gamma \quad (26.7\ \text{MeV}),$$

where the 26.7 MeV includes the annihilation energy of the positrons and electrons and is distributed among all the reaction products. The solar interior is dense, and the reactions occur deep in the Sun where temperatures are highest. It takes about 32,000 years for the energy to diffuse to the surface and radiate away. However, the neutrinos can carry their energy out of the Sun in less than two seconds, because they interact so weakly with other matter. Negative feedback in the Sun acts as a thermostat to regulate the overall energy output. For instance, if the interior of the Sun becomes hotter than normal, the reaction rate increases, producing energy that expands the interior. The expansion cools it and lowers the reaction rate. Conversely, if the interior becomes too cool, it contracts, increasing the temperature and therefore the reaction rate (see Figure 22.31). Stars like the Sun are stable for billions of years, until a significant fraction of their hydrogen has been depleted.

Figure 22.31 Nuclear fusion in the Sun converts hydrogen nuclei into helium; fusion occurs primarily at the boundary of the helium core, where the temperature is highest and sufficient hydrogen remains. Energy released diffuses slowly to the surface, with the exception of neutrinos, which escape immediately. Energy production remains stable because of negative-feedback effects.

Nuclear Weapons and Nuclear Power

The world was in political turmoil when fission was discovered in 1938. Compounding the troubles, the possibility of a self-sustained chain reaction was immediately recognized by leading scientists the world over. The enormous energy known to be in nuclei, but considered inaccessible, now seemed to be available on a large scale.

Within months after the announcement of the discovery of fission, Adolf Hitler banned the export of uranium from newly occupied Czechoslovakia. It seemed that the possible military value of uranium had been recognized in Nazi Germany, and that a serious effort to build a nuclear bomb had begun.

Alarmed scientists, many of whom fled Nazi Germany, decided to take action. None was more famous or revered than Einstein. It was felt that his help was needed to get the American government to make a serious effort at constructing nuclear weapons as a matter of survival. Leo Szilard, a Hungarian physicist who had emigrated to America, took a draft of a letter to Einstein, who, although a pacifist, signed the final version. The letter was for President Franklin Roosevelt, warning of the German potential to build extremely powerful bombs of a new type. It was sent in August of 1939, just before the German invasion of Poland that marked the start of World War II.

It was not until December 6, 1941, the day before the Japanese attack on Pearl Harbor, that the United States made a massive commitment to building a nuclear bomb. The top secret Manhattan Project was a crash program aimed at beating the Germans.

It was carried out in remote locations, such as Los Alamos, New Mexico, whenever possible, and eventually came to cost billions of dollars and employ the efforts of more than 100,000 people. J. Robert Oppenheimer (1904–1967), a talented physicist, was chosen to head the project. The first major step was made by Enrico Fermi and his group in December 1942, when they completed the first self-sustaining nuclear reactor. This first *atomic pile*, built in a squash court at the University of Chicago, proved that a fission chain reaction was possible.

Plutonium was recognized as easier to fission with neutrons and, hence, a superior fission material very early in the Manhattan Project. Plutonium availability was uncertain, and so a uranium bomb was developed simultaneously. Figure 22.32 shows a gun-type bomb, which takes two subcritical uranium masses and shoots them together. To get an appreciable yield, the critical mass must be held together by the explosive charges inside the cannon barrel for a few microseconds. Since the buildup of the uranium chain reaction is relatively slow, the device to bring the critical mass together can be relatively simple. Owing to the fact that the rate of spontaneous fission is low, a neutron source is at the center the assembled critical mass.

Figure 22.32 A gun-type fission bomb for ^{235}U utilizes two subcritical masses forced together by explosive charges inside a cannon barrel. The energy yield depends on the amount of uranium and the time it can be held together before it disassembles itself.

Plutonium's special properties necessitated a more sophisticated critical mass assembly, shown schematically in Figure 22.33. A spherical mass of plutonium is surrounded by shaped charges (high explosives that focus their blast) that implode the plutonium, crushing it into a smaller volume to form a critical mass. The implosion technique is faster and more effective, because it compresses three-dimensionally rather than one-dimensionally as in the gun-type bomb. Again, a neutron source is included to initiate the chain reaction.

Figure 22.33 An implosion created by high explosives compresses a sphere of ^{239}Pu into a critical mass. The superior fissionability of plutonium has made it the preferred bomb material.

Owing to its complexity, the plutonium bomb needed to be tested before there could be any attempt to use it. On July 16, 1945, the test named Trinity was conducted in the isolated Alamogordo Desert in New Mexico, about 200 miles south of Los Alamos (see Figure 22.34). A new age had begun. The yield of the Trinity device was about 10 kilotons (kT), the equivalent of 5,000 of the largest conventional bombs.

Figure 22.34 Trinity test (1945), the first nuclear bomb (credit: U.S. Department of Energy)

Although Germany surrendered on May 7, 1945, Japan had been steadfastly refusing to surrender for many months, resulting large numbers of civilian and military casualties. Invasion plans by the Allies estimated a million casualties of their own and untold losses of Japanese lives. The bomb was viewed as a way to end the war. The first bomb used was a gun-type uranium bomb dropped on Hiroshima on August 6 by the United States. Its yield of about 15 kT destroyed the city and killed an estimated 80,000 people, with 100,000 more being seriously injured. The second bomb was an implosion-type plutonium bomb dropped on Nagasaki only three days later. Its 20-kT yield killed at least 50,000 people, something less than Hiroshima because of the hilly terrain and the fact that it was a few kilometers off target. The Japanese were told that one bomb a week would be dropped until they surrendered unconditionally, which they did on August 14. In actuality, the United States had only enough plutonium for one more bomb, as yet unassembled.

Knowing that fusion produces several times more energy per kilogram of fuel than fission, some scientists pursued the idea of constructing a fusion bomb. The first such bomb was detonated by the United States several years after the first fission bombs, on October 31, 1952, at Eniwetok Atoll in the Pacific Ocean. It had a yield of 10 megatons (MT), about 670 times that of the fission bomb that destroyed Hiroshima. The Soviet Union followed with a fusion device of its own in August 1953, and a weapons race, beyond the aim of this text to discuss, continued until the end of the Cold War.

Figure 22.35 shows a simple diagram of how a thermonuclear bomb is constructed. A fission bomb is exploded next to fusion fuel in the solid form of lithium deuteride. Before the shock wave blows it apart, γ rays heat and compress the fuel, and neutrons create tritium through the reaction $n + {}^{6}\text{Li} \rightarrow {}^{3}\text{H} + {}^{4}\text{He}$. Additional fusion and fission fuels are enclosed in a dense shell of ${}^{238}\text{U}$. At the same time that the uranium shell reflects the neutrons back into the fuel to enhance its fusion, the fast-moving neutrons cause the plentiful and inexpensive ${}^{238}\text{U}$ to fission, part of what allows thermonuclear bombs to be so large.

Figure 22.35 This schematic of a fusion bomb (H-bomb) gives some idea of how the ^{239}Pu fission trigger is used to ignite fusion fuel. Neutrons and γ rays transmit energy to the fusion fuel, create tritium from deuterium, and heat and compress the fusion fuel. The outer shell of ^{238}U serves to reflect some neutrons back into the fuel, causing more fusion, and it boosts the energy output by fissioning itself when neutron energies become high enough.

Of course, not all applications of nuclear physics are as destructive as the weapons described above. Hundreds of nuclear fission power plants around the world attest to the fact that controlled fission is both practical and economical. Given growing concerns over global warming, nuclear power is often seen as a viable alternative to energy derived from fossil fuels.

BOUNDLESS PHYSICS

Fusion Reactors

For decades, fusion reactors have been deemed the *energy of the future*. A safer, cleaner, and more abundant potential source of energy than its fission counterpart, images of the fusion reactor have been conjured up each time the need for a renewable, environmentally friendly resource is discussed. Now, after more than half a century of speculating, some scientists believe that fusion reactors are nearly here.

In creating energy by combining atomic nuclei, the fusion reaction holds many advantages over fission. First, fusion reactions are more efficient, releasing 3 to 4 times more energy than fission per gram of fuel. Furthermore, unlike fission reactions that require *heavy* elements like uranium that are difficult to obtain, fusion requires *light* elements that are abundant in nature. The greatest advantage of the fusion reaction, however, is in its ability to be controlled. While traditional nuclear reactors create worries about meltdowns and radioactive waste, neither is a substantial concern with the fusion reaction. Consider that fusion reactions require a large amount of energy to overcome the repulsive Coulomb force and that the byproducts of a fusion reaction are largely limited to helium nuclei.

In order for fusion to occur, hydrogen isotopes of deuterium and tritium must be acquired. While deuterium can easily be gathered from ocean water, tritium is slightly more difficult to come by, though it can be manufactured from Earth's abundant lithium. Once acquired, the hydrogen isotopes are injected into an empty vessel and subjected to temperature and pressure great enough to mimic the conditions at the core of our Sun. Using carefully controlled high-frequency radio waves, the hydrogen isotopes are broken into plasma and further controlled through an electromagnetic field. As the electromagnetic field continues to exert pressure on the hydrogen plasma, enough energy is supplied to cause the hydrogen plasma to fuse into helium.

Figure 22.36 Tokamak confinement of nuclear fusion plasma. The magnetic field lines are used to confine the high-temperature plasma (purple). Research is currently being done to increase the efficiency of the tokamak confinement model.

Once the plasma fuses, high-velocity neutrons are ejected from the newly formed helium atoms. Those high velocity neutrons, carrying the excess energy stored within bonds of the original hydrogen, are able to travel unaffected by the applied magnetic field. In doing so, they strike a barrier around the nuclear reactor, transforming their excess energy to heat. The heat is then harvested to make steam that drives turbines. Hydrogen's tremendous power is now usable!

The historical concern with nuclear fusion reactors is that the energy required to control the electromagnetic field is greater than the energy harvested from the hydrogen atoms. However, recent research by both Lockheed Martin engineers and scientists at the Lawrence Livermore National Laboratory has yielded exciting theoretical improvements in efficiency. At the time of this writing, a test facility called ITER (International Thermonuclear Experimental Reactor) is being constructed in southern France. A joint venture of the European Union, the United States, Japan, Russia, China, South Korea, and India, ITER is designed for further study into the future of nuclear fusion energy production.

22.5 Medical Applications of Radioactivity: Diagnostic Imaging and Radiation

Section Learning Objectives

By the end of this section, you will be able to do the following:

- Describe how nuclear imaging works (e.g., radioisotope imaging, PET)
- Describe the ionizing effects of radiation and how they can be used for medical treatment

Section Key Terms

Anger camera	rad	radiopharmaceutical	therapeutic ratio
relative biological effectiveness (RBE)	roentgen equivalent man (rem)	tagged	

Medical Applications of Nuclear Physics

Applications of nuclear physics have become an integral part of modern life. From the bone scan that detects one cancer to the radioiodine treatment that cures another, nuclear radiation has diagnostic and therapeutic effects on medicine.

Medical Imaging

A host of medical imaging techniques employ nuclear radiation. What makes nuclear radiation so useful? First, γ radiation can

easily penetrate tissue; hence, it is a useful probe to monitor conditions inside the body. Second, nuclear radiation depends on the nuclide and not on the chemical compound it is in, so that a radioactive nuclide can be put into a compound designed for specific purposes. When that is done, the compound is said to be **tagged**. A tagged compound used for medical purposes is called a **radiopharmaceutical**. Radiation detectors external to the body can determine the location and concentration of a radiopharmaceutical to yield medically useful information. For example, certain drugs are concentrated in inflamed regions of the body, and their locations can aid diagnosis and treatment as seen in Figure 22.37. Another application utilizes a radiopharmaceutical that the body sends to bone cells, particularly those that are most active, to detect cancerous tumors or healing points. Images can then be produced of such bone scans. Clever use of radioisotopes determines the functioning of body organs, such as blood flow, heart muscle activity, and iodine uptake in the thyroid gland. For instance, a radioactive form of iodine can be used to monitor the thyroid, a radioactive thallium salt can be used to follow the blood stream, and radioactive gallium can be used for cancer imaging.

Figure 22.37 A radiopharmaceutical was used to produce this brain image of a patient with Alzheimer's disease. Certain features are computer enhanced. (credit: National Institutes of Health)

Once a radioactive compound has been ingested, a device like that shown in Figure 22.38 is used to monitor nuclear activity. The device, called an **Anger camera** or gamma camera uses a piece of lead with holes bored through it. The gamma rays are redirected through the collimator to narrow their beam, and are then interpreted using a device called a scintillator. The computer analysis of detector signals produces an image. One of the disadvantages of this detection method is that there is no depth information (i.e., it provides a two-dimensional view of the tumor as opposed to a three-dimensional view), because radiation from any location under that detector produces a signal.

Figure 22.38 An Anger or gamma camera consists of a lead collimator and an array of detectors. Gamma rays produce light flashes in the scintillators. The light output is converted to an electrical signal by the photomultipliers. A computer constructs an image from the detector output.

Single-photon-emission computer tomography (SPECT) used in conjunction with a CT scanner improves on the process carried out by the gamma camera. Figure 22.39 shows a patient in a circular array of SPECT detectors that may be stationary or rotated, with detector output used by a computer to construct a detailed image. The spatial resolution of this technique is poor, but the three-dimensional image created results in a marked improvement in contrast.

Figure 22.39 SPECT uses a rotating camera to form an image of the concentration of a radiopharmaceutical compound. (credit: Woldo, Wikimedia Commons)

Positron emission tomography (or PET) scans utilize images produced by β^+ emitters. When the emitted positron β^+ encounters an electron, mutual annihilation occurs, producing two γ rays. Those γ rays have identical 0.511 MeV energies (the energy comes from the destruction of an electron or positron mass) and they move directly away from each other, allowing detectors to determine their point of origin accurately (as shown in Figure 22.40). It requires detectors on opposite sides to simultaneously (i.e., at the same time) detect photons of 0.511 MeV energy and utilizes computer imaging techniques similar to those in SPECT and CT scans. PET is used extensively for diagnosing brain disorders. It can note decreased metabolism in certain regions that accompany Alzheimer's disease. PET can also locate regions in the brain that become active when a person carries out specific activities, such as speaking, closing his or her eyes, and so on.

Figure 22.40 A PET system takes advantage of the two identical γ -ray photons produced by positron-electron annihilation. The γ rays are emitted in opposite directions, so that the line along which each pair is emitted is determined. Various events detected by several pairs of detectors are then analyzed by the computer to form an accurate image.

Ionizing Radiation on the Body

We hear many seemingly contradictory things about the biological effects of ionizing radiation. It can cause cancer, burns, and hair loss, and yet it is used to treat and even cure cancer. How do we understand such effects? Once again, there is an underlying simplicity in nature, even in complicated biological organisms. All the effects of ionizing radiation on biological tissue can be understood by knowing that ionizing radiation affects molecules within cells, particularly DNA molecules. Let us take a brief look at molecules within cells and how cells operate. Cells have long, double-helical DNA molecules containing chemical patterns called genetic codes that govern the function and processes undertaken by the cells. Damage to DNA consists of breaks in chemical bonds or other changes in the structural features of the DNA chain, leading to changes in the genetic code. In human cells, we can have as many as a million individual instances of damage to DNA per cell per day. The repair ability of DNA is vital for maintaining the integrity of the genetic code and for the normal functioning of the entire organism. A cell with a damaged ability to repair DNA, which could have been induced by ionizing radiation, can do one of the following:

- The cell can go into an irreversible state of dormancy, known as senescence.
- The cell can commit suicide, known as programmed cell death.
- The cell can go into unregulated cell division, leading to tumors and cancers.

Since ionizing radiation damages the DNA, ionizing radiation has its greatest effect on cells that rapidly reproduce, including most types of cancer. Thus, cancer cells are more sensitive to radiation than normal cells and can be killed by it easily. Cancer is characterized by a malfunction of cell reproduction, and can also be caused by ionizing radiation. There is no contradiction to say that ionizing radiation can be both a cure and a cause.

Radiotherapy

Radiotherapy is effective against cancer because cancer cells reproduce rapidly and, consequently, are more sensitive to radiation. The central problem in radiotherapy is to make the dose for cancer cells as high as possible while limiting the dose for normal cells. The ratio of abnormal cells killed to normal cells killed is called the **therapeutic ratio**, and all radiotherapy techniques are designed to enhance that ratio. Radiation can be concentrated in cancerous tissue by a number of techniques. One of the most prevalent techniques for well-defined tumors is a geometric technique shown in Figure 22.41. A narrow beam of radiation is passed through the patient from a variety of directions with a common crossing point in the tumor. The technique concentrates the dose in the tumor while spreading it out over a large volume of normal tissue.

Figure 22.41 The ^{60}Co source of γ -radiation is rotated around the patient so that the common crossing point is in the tumor, concentrating the dose there. This geometric technique works for well-defined tumors.

Another use of radiation therapy is through radiopharmaceuticals. Cleverly, radiopharmaceuticals are used in cancer therapy by tagging antibodies with radioisotopes. Those antibodies are extracted from the patient, cultured, loaded with a radioisotope, and then returned to the patient. The antibodies are then concentrated almost entirely in the tissue they developed to fight, thus localizing the radiation in abnormal tissue. This method is used with radioactive iodine to fight thyroid cancer. While the therapeutic ratio can be quite high for such short-range radiation, there can be a significant dose for organs that eliminate radiopharmaceuticals from the body, such as the liver, kidneys, and bladder. As with most radiotherapy, the technique is limited by the tolerable amount of damage to the normal tissue.

Radiation Dosage

To quantitatively discuss the biological effects of ionizing radiation, we need a radiation dose unit that is directly related to those effects. To do define such a unit, it is important to consider both the biological organism and the radiation itself. Knowing that the amount of ionization is proportional to the amount of deposited energy, we define a radiation dose unit called the **rad**. It 1/100 of a joule of ionizing energy deposited per kilogram of tissue, which is

$$1 \text{ rad} = 0.01 \text{ J/kg}. \tag{22.69}$$

For example, if a 50.0-kg person is exposed to ionizing radiation over her entire body and she absorbs 1.00 J, then her whole-body radiation dose is

$$(1.00 \text{ J})/(50.0 \text{ kg}) = 0.0200 \text{ J/kg} = 2.00 \text{ rad}. \tag{22.70}$$

If the same 1.00 J of ionizing energy were absorbed in her 2.00-kg forearm alone, then the dose to the forearm would be

$$(1.00 \text{ J})/(2.00 \text{ kg}) = 0.500 \text{ J/kg} = 50.0 \text{ rad}, \tag{22.71}$$

and the unaffected tissue would have a zero rad dose. When calculating radiation doses, you divide the energy absorbed by the mass of affected tissue. You must specify the affected region, such as the whole body or forearm in addition to giving the numerical dose in rads. Although the energy per kilogram in 1 rad is small, it can still have significant effects. Since only a few eV cause ionization, just 0.01 J of ionizing energy can create a huge number of ion pairs and have an effect at the cellular level.

The effects of ionizing radiation may be directly proportional to the dose in rads, but they also depend on the type of radiation and the type of tissue. That is, for a given dose in rads, the effects depend on whether the radiation is α, β, γ, X-ray, or some

other type of ionizing radiation. The **relative biological effectiveness** (RBE) relates to the amount of biological damage that can occur from a given type of radiation and is given in Table 22.4 for several types of ionizing radiation.

Type and energy of radiation	RBE
X-rays	1
γ rays	1
β rays greater than 32 keV	1
β rays less than 32 keV	1.7
Neutrons, thermal to slow (< 20 keV)	2–5
Neutrons, fast (1–10 MeV)	10 (body), 32 (eyes)
Protons (1–10 MeV)	10 (body), 32 (eyes)
α rays from radioactive decay	10–20
Heavy ions from accelerators	10–20

Table 22.4 Relative Biological Effectiveness

TIPS FOR SUCCESS

The RBEs given in Table 22.4 are approximate, but they yield certain valuable insights.

- The eyes are more sensitive to radiation, because the cells of the lens do not repair themselves.
- Though both are neutral and have large ranges, neutrons cause more damage than γ rays because neutrons often cause secondary radiation when they are captured.
- Short-range particles such as α rays have a severely damaging effect to internal anatomy, as their damage is concentrated and more difficult for the biological organism to repair. However, the skin can usually block alpha particles from entering the body.

Can you think of any other insights from the table?

A final dose unit more closely related to the effect of radiation on biological tissue is called the **roentgen equivalent man**, or rem. A combination of all factors mentioned previously, the roentgen equivalent man is defined to be the dose in rads multiplied by the relative biological effectiveness.

$$\text{rem} = \text{rad} \times \text{RBE} \tag{22.72}$$

The large-scale effects of radiation on humans can be divided into two categories: immediate effects and long-term effects. Table 22.5 gives the immediate effects of whole-body exposures received in less than one day. If the radiation exposure is spread out over more time, greater doses are needed to cause the effects listed. Any dose less than 10 rem is called a low dose, a dose 10 to 100 rem is called a moderate dose, and anything greater than 100 rem is called a high dose.

Dose (rem)	Effect
0–10	No observable effect
10–100	Slight to moderate decrease in white blood cell counts

Table 22.5 Immediate Effects of Radiation (Adults, Whole Body, Single Exposure)

Dose (rem)	Effect
50	Temporary sterility
100–200	Significant reduction in blood cell counts, brief nausea, and vomiting; rarely fatal
200–500	Nausea, vomiting, hair loss, severe blood damage, hemorrhage, fatalities
450	LD50/32; lethal to 50% of the population within 32 days after exposure if untreated
500–2,000	Worst effects due to malfunction of small intestine and blood systems; limited survival
> 2,000	Fatal within hours due to collapse of central nervous system

Table 22.5 Immediate Effects of Radiation (Adults, Whole Body, Single Exposure)

WORK IN PHYSICS

Health Physicist

Are you interested in learning more about radiation? Are you curious about studying radiation dosage levels and ensuring the safety of the environment and people that are most closely affected by it? If so, you may be interested in becoming a health physicist.

The field of health physics draws from a variety of science disciplines with the central aim of mitigating radiation concerns. Those that work as health physicists have a diverse array of potential jobs available to them, including those in research, industry, education, environmental protection, and governmental regulation. Furthermore, while the term *health physicist* may lead many to think of the medical field, there are plenty of applications within the military, industrial, and energy fields as well.

As a researcher, a health physicist can further environmental studies on the effects of radiation, design instruments for more accurate measurements, and assist in establishing valuable radiation standards. Within the energy field, a health physicsist often acts as a manager, closely tied to all operations at all levels, from procuring appropirate equipment to monitoring health data. Within industry, the health physicist acts as a consultant, assisting industry management in important decisions, designing facilities, and choosing appropriate detection tools. The health physicist possesses a unique knowledge base that allows him or her to operate in a wide variety of interesting disciplines!

To become a health physicist, it is necessary to have a background in the physical sciences. Understanding the fields of biology, physiology, biochemistry, and genetics are all important as well. The ability to analyze and solve new problems is critical, and a natural aptitude for science and mathematics will assist in the continued necessary training. There are two possible certifications for health physicists: from the American Board of Health Physicists (ABHP) and the National Registry of Radiation Protection Technologists (NRRPT).

KEY TERMS

activity rate of decay for radioactive nuclides
alpha decay type of radioactive decay in which an atomic nucleus emits an alpha particle
anger camera common medical imaging device that uses a scintillator connected to a series of photomultipliers
atomic number number of protons in a nucleus
becquerel SI unit for rate of decay of a radioactive material
beta decay type of radioactive decay in which an atomic nucleus emits a beta particle
carbon-14 dating radioactive dating technique based on the radioactivity of carbon-14
chain reaction self-sustaining sequence of events, exemplified by the self-sustaining nature of a fission reaction at critical mass
critical mass minimum amount necessary for self-sustained fission of a given nuclide
decay constant quantity that is inversely proportional to the half-life and that is used in the equation for number of nuclei as a function of time
energy-level diagram a diagram used to analyze the energy levels of electrons in the orbits of an atom
excited state any state beyond the $n = 1$ orbital in which the electron stores energy
Fraunhofer lines black lines shown on an absorption spectrum that show the wavelengths absorbed by a gas
gamma decay type of radioactive decay in which an atomic nucleus emits a gamma ray
Geiger tube very common radiation detector that usually gives an audio output
ground state the n=1 orbital of an electron
half-life time in which there is a 50 percent chance that a nucleus will decay
Heisenberg uncertainty principle fundamental limit to the precision with which pairs of quantities such as momentum and position can be measured
hydrogen-like atom any atom with only a single electron
isotope nuclei having the same Z and different N's
liquid drop model model of the atomic nucleus (useful only to understand some of its features) in which nucleons in a nucleus act like atoms in a drop
mass number number of nucleons in a nucleus
nuclear fission reaction in which a nucleus splits
nuclear fusion reaction in which two nuclei are combined, or fused, to form a larger nucleus
nucleons particles found inside nuclei
planetary model of the atom model of the atom that shows electrons orbiting like planets about a Sun-like nucleus
proton-proton cycle combined reactions
$^1\text{H} + ^1\text{H} \rightarrow {}^2\text{H} + e^- + \nu_e$.
$^1\text{H} + ^2\text{H} \rightarrow {}^3\text{He} + \gamma$ and
$^3\text{He} + ^3\text{He} \rightarrow {}^4\text{He} + ^1\text{H} + ^1\text{H}$ that begins with hydrogen and ends with helium
rad amount of ionizing energy deposited per kilogram of tissue
radioactive substance or object that emits nuclear radiation
radioactive dating application of radioactive decay in which the age of a material is determined by the amount of radioactivity of a particular type that occurs
radioactive decay process by which an atomic nucleus of an unstable atom loses mass and energy by emitting ionizing particles
radioactivity emission of rays from the nuclei of atoms
radiopharmaceutical compound used for medical imaging
relative biological effectiveness (RBE) number that expresses the relative amount of damage that a fixed amount of ionizing radiation of a given type can inflict on biological tissues
roentgen equivalent man (rem) dose unit more closely related to effects in biological tissue
Rutherford scattering scattering of alpha particles by gold nuclei in the gold foil experiment
Rydberg constant a physical constant related to atomic spectra, with an established value of $1.097 \times 10^7 \text{m}^{-1}$
scintillator radiation detection method that records light produced when radiation interacts with materials
strong nuclear force attractive force that holds nucleons together within the nucleus
tagged having a radioactive substance attached (to a chemical compound)
therapeutic ratio the ratio of abnormal cells killed to normal cells killed
transmutation process of changing elemental composition

SECTION SUMMARY

22.1 The Structure of the Atom

- Rutherford's gold foil experiment provided evidence that the atom is composed of a small, dense nucleus with electrons occupying the mostly empty space around it.
- Analysis of emission spectra shows that energy is emitted from energized gas in discrete quantities.
- The Bohr model of the atom describes electrons existing in discrete orbits, with discrete energies emitted and absorbed as the electrons decrease and increase in orbital energy.
- The energy emitted or absorbed by an electron as it changes energy state can be determined with the equation $\Delta E = E_i - E_f$, where

$E_n = \frac{Z^2}{n^2} E_o \ (n = 1, 2, 3, ...)$.

- The wavelength of energy absorbed or emitted by an electron as it changes energy state can be determined by the equation $\frac{1}{\lambda} = R\left(\frac{1}{n_f{}^2} - \frac{1}{n_i{}^2}\right)$, where $R = 1.097 \times 10^7 \text{m}^{-1}$.
- Described as an electron cloud, the quantum model of the atom is the result of de Broglie waves and Heisenberg's uncertainty principle.

22.2 Nuclear Forces and Radioactivity

- The structure of the nucleus is defined by its two nucleons, the neutron and proton.
- Atomic numbers and mass numbers are used to differentiate between various atoms and isotopes. Those numbers can be combined into an easily recognizable form called a *nuclide*.
- The size and stability of the nucleus is based upon two forces: the electromagnetic force and strong nuclear force.
- Radioactive decay is the alteration of the nucleus through the emission of particles or energy.
- Alpha decay occurs when too many protons exist in the nucleus. It results in the ejection of an alpha particle, as described in the equation ${}^A_Z X_N \rightarrow {}^{A-4}_{Z-2} Y_N + {}^4_2\text{He}$.
- Beta decay occurs when too many neutrons (or protons) exist in the nucleus. It results in the transmutation of a neutron into a proton, electron, and neutrino. The decay is expressed through the equation ${}^A_Z X_N \rightarrow {}_{Z+1}^{\ \ A} Y_{N-1} + e + \nu$. (Beta decay may also transform a proton into a neutron.)
- Gamma decay occurs when a nucleus in an excited state move to a more stable state, resulting in the release of a photon. Gamma decay is represented with the equation ${}^A_Z X_N \rightarrow X_N + \gamma$.
- The penetration distance of radiation depends on its energy, charge, and type of material it encounters.

22.3 Half Life and Radiometric Dating

- Radioactive half-life is the time it takes a sample of nuclei to decay to half of its original amount.
- The rate of radioactive decay is defined as the sample's activity, represented by the equation $R = \frac{\Delta N}{\Delta t}$.
- Knowing the half-life of a radioactive isotope allows for the process of radioactive dating to determine the age of a material.
- If the half-life of a material is known, the age of the material can be found using the equation $N = N_O e^{-\lambda t}$.
- The age of organic material can be determined using the decay of the carbon-14 isotope, while the age of rocks can be determined using the decay of uranium-238.

22.4 Nuclear Fission and Fusion

- Nuclear fission is the splitting of an atomic bond, releasing a large amount of potential energy previously holding the atom together. The amount of energy released can be determined through the equation $E = mc^2$.
- Nuclear fusion is the combining, or fusing together, of two nuclei. Energy is also released in nuclear fusion as the combined nuclei are closer together, resulting in a decreased strong nuclear force.
- Fission was used in two nuclear weapons at the conclusion of World War II: the gun-type uranium bomb and the implosion-type plutonium bomb.
- While fission has been used in both nuclear weapons and nuclear reactors, fusion is capable of releasing more energy per reaction. As a result, fusion is a well-researched, if not yet well-controlled, energy source.

22.5 Medical Applications of Radioactivity: Diagnostic Imaging and Radiation

- Medical imaging occurs when a radiopharmaceutical placed in the body provides information to an array of radiation detectors outside the body.
- Devices utilizing medical imaging include the Anger camera, SPECT detector, and PET scan.
- Ionizing radiation can both cure and cause cancer through the manipulation of DNA molecules.
- Radiation dosage and its effect on the body can be measured using the quantities radiation dose unit (rad), relative biological effectiveness (RBE), and the roentgen equivalent man (rem).

KEY EQUATIONS

22.1 The Structure of the Atom

energy of hydrogen electron in an orbital	$E_n = -\frac{13.6 \text{ eV}}{n^2} (n = 1, 2, 3, ...)$
energy of any hydrogen-like electron in orbital	$E_n = \frac{Z^2}{n^2} E_o \ (n = 1, 2, 3, ...)$

wavelength of light emitted by an electron changing states	$\frac{1}{\lambda} = R\left(\frac{1}{{n_f}^2} - \frac{1}{{n_i}^2}\right)$
wavelength of an orbital	$n\lambda_n = 2\pi r_n \ (n = 1, 2, 3, ...)$
heisenberg's uncertainty principle	$\Delta x \Delta p \geq \frac{h}{4\pi}$

22.2 Nuclear Forces and Radioactivity

alpha decay equation	${}^{A}_{Z}X_N \rightarrow {}^{A-4}_{Z-2}Y_N + {}^{4}_{2}\text{He}$
beta decay equation	${}^{A}_{Z}X_N \rightarrow {}^{A}_{Z+1}Y_{N-1} + e + \nu$
gamma decay equation	${}^{A}_{Z}X_N \rightarrow X_N + \gamma$

22.3 Half Life and Radiometric Dating

radioactive half-life	$N = N_O e^{-\lambda t}$

22.4 Nuclear Fission and Fusion

energy–mass conversion	$E = mc^2$
proton-proton chain	$2e- + 4\,{}^{1}\text{H} \rightarrow 4\ \text{He} + 2\nu_e + 6\gamma$

22.5 Medical Applications of Radioactivity: Diagnostic Imaging and Radiation

roentgen equivalent man	$\text{rem} = \text{rad} \times \text{RBE}$

CHAPTER REVIEW

Concept Items

22.1 The Structure of the Atom

1. A star emits light from its core. One observer views the emission unobstructed while a second observer views the emission while obstructed by a cloud of hydrogen gas. Describe the difference between their observations.
   a. Intensity of the light in the spectrum will increase.
   b. Intensity of the light in the spectrum will decrease.
   c. Frequencies will be absorbed from the spectrum.
   d. Frequencies will be added to the spectrum.

2. How does the orbital energy of a hydrogen-like atom change as it increases in atomic number?
   a. The orbital energy will increase.
   b. The orbital energy will decrease.
   c. The orbital energy will remain constant.
   d. The orbital energy will be halved.

22.4 Nuclear Fission and Fusion

3. Aside from energy yield, why are nuclear fusion reactors more desirable than nuclear fission reactors?
   a. Nuclear fusion reactors have a low installation cost.
   b. Radioactive waste is greater for a fusion reactor.
   c. Nuclear fusion reactors are easy to design and build.
   d. A fusion reactor produces less radioactive waste.

Critical Thinking Items

22.1 The Structure of the Atom

4. How would the gold foil experiment have changed if electrons were used in place of alpha particles, assuming that the electrons hit the gold foil with the same force as the alpha particles?
   a. Being less massive, the electrons might have been scattered to a greater degree than the alpha particles.
   b. Being less massive, the electrons might have been scattered to a lesser degree than the alpha particles.
   c. Being more massive, the electrons would have been scattered to a greater degree than the alpha particles.
   d. Being more massive, the electrons would have been scattered to a lesser degree than the alpha particles.

5. Why does the emission spectrum of an isolated gas differ from the emission spectrum created by a white light?
   a. White light and an emission spectrum are different varieties of continuous distribution of frequencies.
   b. White light and an emission spectrum are different series of discrete frequencies.

c. White light is a continuous distribution of frequencies, and an emission spectrum is a series of discrete frequencies.
d. White light is a series of discrete frequencies, and an emission spectrum is a continuous distribution of frequencies.

6. Why would it most likely be difficult to observe quantized orbital states for satellites orbiting the earth?
a. On a macroscopic level, the orbital states do exist for satellites orbiting Earth but are too closely spaced for us to see.
b. On a macroscopic level, the orbital states do not exist for satellites orbiting Earth.
c. On a macroscopic level, we cannot control the amount of energy that we give to an artificial satellite and thus control its orbital altitude.
d. On a macroscopic level, we cannot control the amount of energy that we give to an artificial satellite but we can control its orbital altitude.

7. Do standing waves explain why electron orbitals are quantized?
a. no
b. yes

8. Some terms referring to the observation of light include *emission spectrum* and *absorption spectrum*. Based on these definitions, what would a *reflection spectrum* describe?
a. The reflection spectrum would describe when incident waves are selectively reflected by a substance.
b. The reflection spectrum would describe when incident waves are completely reflected by a substance.
c. The reflection spectrum would describe when incident waves are not absorbed by a substance.
d. The reflection spectrum would describe when incident waves are completely absorbed by a substance.

22.2 Nuclear Forces and Radioactivity

9. Explain why an alpha particle can have a greater range in air than a beta particle in lead.
a. While the alpha particle has a lesser charge than a beta particle, the electron density in lead is much less than that in air.
b. While the alpha particle has a greater charge than a beta particle, the electron density in lead is much lower than that in air.
c. While the alpha particle has a lesser charge than a beta particle, the electron density in lead is much greater than that in air.
d. While the alpha particle has a greater charge than a beta particle, the electron density in lead is much higher than that in air.

10. What influence does the strong nuclear force have on the electrons in an atom?
a. It attracts them toward the nucleus.
b. It repels them away from the nucleus.
c. The strong force makes electrons revolve around the nucleus.
d. It does not have any influence.

22.3 Half Life and Radiometric Dating

11. Provide an example of something that decreases in a manner similar to radioactive decay.
a. The potential energy of an object falling under the influence of gravity
b. The kinetic energy of a ball that is dropped from a building to the ground
c. Theh charge transfer from an ebonite rod to fur
d. The heat transfer from a hot to a cold object

12. A sample of radioactive material has a decay constant of $0.05\ s^{-1}$. Why is it wrong to presume that the sample will take just 20 seconds to fully decay?
a. The decay constant varies with the mass of the sample.
b. The decay constant results vary with the amount of the sample.
c. The decay constant represents a percentage of the sample that cannot decay.
d. The decay constant represents only the fraction of a sample that decays in a unit of time, not the decay of the entire sample.

22.4 Nuclear Fission and Fusion

13. What is the atomic number of the most strongly bound nuclide?
a. 25
b. 26
c. 27
d. 28

14. Why are large electromagnets necessary in nuclear fusion reactors?
a. Electromagnets are used to slow down the movement of charge hydrogen plasma.
b. Electromagnets are used to decrease the temperature of hydrogen plasma.
c. Electromagnets are used to confine the hydrogen plasma.
d. Electromagnets are used to stabilize the temperature of the hydrogen plasma.

22.5 Medical Applications of Radioactivity: Diagnostic Imaging and Radiation

15. Why are different radiopharmaceuticals used to image different parts of the body?
   a. The different radiopharmaceuticals travel through different blood vessels.
   b. The different radiopharmaceuticals travel to different parts of the body.
   c. The different radiopharmaceuticals are used to treat different diseases of the body.
   d. The different radiopharmaceuticals produce different amounts of ionizing radiation.

16. Why do people think carefully about whether to receive a diagnostic test such as a CT scan?
   a. The radiation from a CT scan is capable of creating cancerous cells.
   b. The radiation from a CT scan is capable of destroying cancerous cells.
   c. The radiation from a CT scan is capable of creating diabetic cells.
   d. The radiation from a CT scan is capable of destroying diabetic cells.

17. Sometimes it is necessary to take a PET scan very soon after ingesting a radiopharmaceutical. Why is that the case?
   a. The radiopharmaceutical may have a short half-life.
   b. The radiopharmaceutical may have a long half-life.
   c. The radiopharmaceutical quickly passes through the digestive system.
   d. The radiopharmaceutical can become lodged in the digestive system.

Performance Task

22.5 Medical Applications of Radioactivity: Diagnostic Imaging and Radiation

18. On the Environmental Protection Agency's website, a helpful tool exists to allow you to determine your average annual radiation dose. Use the tool to determine whether the radiation level you have been exposed to is dangerous and to compare your radiation dosage to other radiative events.

   1. Visit the webpage (http://www.openstax.org/l/28calculate) and answer the series of questions provided to determine the average annual radiation dosage that you receive.
   2. Table 22.5 shows the immediate effects of a radiation dosage. Using the table, explain what you would experience if your yearly dosage of radiation was received all over the course of one day. Also, determine whether your dosage is considered a low, moderate, or high.
   3. Using the information input into the webpage, what percentage of your dosage comes from natural sources? The average percentage of radiation from natural sources for an individual is around 85 percent.
   4. Research radiation dosages for evacuees from events like the Chernobyl and Fukushima meltdowns. How does your annual radiation exposure rate compare to the net dosage for evacuees of each event. Use numbers to support your answer.
   5. The U.S. Department of Labor limits the amount of radiation that a given worker may receive in a 12 month period.
      a. Research the present maximum value and compare your annual exposure rate to that of a radiation worker. Use numbers to support your answer.
      b. What types of work are likely to cause an increase in the radiation exposure of a particular worker?

Provide one question based upon the information gathered on the EPA website.

TEST PREP

Multiple Choice

22.1 The Structure of the Atom

19. If electrons are negatively charged and the nucleus is positively charged, why do they not attract and collide with each other?
   a. The pull from the nucleus provides a centrifugal force, which is not strong enough to draw the electrons into the nucleus.
   b. The pull from the nucleus provides a centripetal force, which is not strong enough to draw the electrons into the nucleus.

c. The pull from the nucleus provides a helical motion.
d. The pull from the nucleus provides a cycloid motion.

22.4 Nuclear Fission and Fusion

20. If a nucleus elongates due to a neutron strike, which of the following forces will decrease?
a. Nuclear force between neutrons only
b. Coulomb force between protons only
c. Strong nuclear force between all nucleons and Coulomb force between protons, but the strong force will decrease more
d. Strong nuclear force between neutrons and Coulomb force between protons, but Coulomb force will decrease more

Short Answer

22.1 The Structure of the Atom

21. Why do Bohr's calculations for electron energies not work for all atoms?
a. In atoms with more than one electron is an atomic shell, the electrons will interact. That requires a more complex formula than Bohr's calculations accounted for.
b. In atoms with 10 or more electorns in an atomic shell, the electrons will interact. That requires a more complex formula than Bohr's calculations accounted for.
c. In atoms with more than one electron in an atomic shell, the electrons will not interact. That requires a more complex formula than Bohr's calculations accounted for.
d. In atoms with 10 or more electrons in an atomic shell, the electrons will not interact. That requires a more complex formula than Bohr's calculations accounted for.

22.2 Nuclear Forces and Radioactivity

22. Does transmutation occur within chemical reactions?
a. no
b. yes

22.3 Half Life and Radiometric Dating

23. How does the radioactive activity of a sample change with time?
a. The radioactive activity decreases exponentially.
b. The radioactive activity undergoes linear decay.
c. The radioactive activity undergoes logarithmic decay.
d. The radioactive activity will not change with time.

22.4 Nuclear Fission and Fusion

24. Why does fission of heavy nuclei result in the release of neutrons?
a. Heavy nuclei require more neutrons to achieve stability.
b. Heavy nuclei require more neutrons to balance charge.
c. Light nuclei require more neutrons to achieve stability.
d. Light nuclei require more neutrons to balance charge.

22.5 Medical Applications of Radioactivity: Diagnostic Imaging and Radiation

25. Why is radioactive iodine used to monitor the thyroid?
a. Radioactive iodine can be used by the thyroid while absorbing information about the thyroid.
b. Radioactive iodine can be used by the thyroid while emitting information about the thyroid.
c. Radioactive iodine can be secreted by the thyroid while absorbing information about the thyroid.
d. Radioactive iodine can be secreted by the thyroid while emitting information about the thyroid.

Extended Response

22.1 The Structure of the Atom

26. Compare the standing wavelength of an $n = 2$ orbital to the standing wavelength of an $n = 4$ orbital.
a. The standing wavelength of an $n = 2$ orbital is greater than the standing wavelength of an $n = 4$ orbital.
b. The standing wavelength of an $n = 2$ orbital is less than the standing wavelength of an $n = 4$ orbital.
c. There is no relation between the standing wavelength of an $n = 2$ orbital and the standing wavelength of an $n = 4$ orbital.
d. The standing wavelength of an $n = 2$ orbital is the same as the standing wavelength of an$n = 4$ orbital.

27. Describe the shape of the electron cloud, based on total energy levels, for an atom with electrons in multiple orbital states.
a. There are multiple regions of high electron

probability of various shapes surrounding the nucleus.

b. There is a single solid spherical region of high electron probability surrounding the nucleus.

c. There are multiple concentric shells of high electron probability surrounding the nucleus.

d. There is a single spherical shell of high electron probability surrounding the nucleus.

22.2 Nuclear Forces and Radioactivity

28. How did Becquerel's observations of pitchblende imply the existence of radioactivity?

a. A chemical reaction occurred on the photographic plate without any external source of energy.

b. Bright spots appeared on the photographic plate due to an external source of energy.

c. Energy from the Sun was absorbed by the pitchblende and reflected onto the photographic plate.

d. Dark spots appeared on the photographic plate due to an external source of energy.

22.4 Nuclear Fission and Fusion

29. Describe the potential energy of two nuclei as they approach each other.

a. The potential energy will decrease as the nuclei are brought together and then rapidly increase once a minimum is reached.

b. The potential energy will decrease as the nuclei are brought together.

c. The potential energy will increase as the nuclei are brought together.

d. The potential energy will increase as the nuclei are brought together and then rapidly decrease once a maximum is reached.

22.5 Medical Applications of Radioactivity: Diagnostic Imaging and Radiation

30. Why do X-rays and gamma rays have equivalent RBE values if they provide different amounts of energy to the body?

a. The penetration distance, which depends on energy, is short for both X-rays and gamma rays.

b. The penetration distance, which depends on energy, is long for both X-rays and gamma rays.

c. The penetration distance, as determined by their high mass, is different for both X-rays and gamma rays.

d. The penetration distance, as determined by their low mass, is the same for both X-rays and gamma rays.

CHAPTER 23
Particle Physics

Figure 23.1 Part of the Large Hadron Collider (LHC) at CERN, on the border of Switzerland and France. The LHC is a particle accelerator, designed to study fundamental particles. (credit: Image Editor, Flickr)

Chapter Outline

23.1 The Four Fundamental Forces

23.2 Quarks

23.3 The Unification of Forces

INTRODUCTION Following ideas remarkably similar to those of the ancient Greeks, we continue to look for smaller and smaller structures in nature, hoping ultimately to find and understand the most fundamental building blocks that exist. Atomic physics deals with the smallest units of elements and compounds. In its study, we have found a relatively small number of atoms with systematic properties, and these properties have explained a tremendous range of phenomena. Nuclear physics is concerned with the nuclei of atoms and their substructures. Here, a smaller number of components—the proton and neutron—make up all nuclei. Exploring the systematic behavior of their interactions has revealed even more about matter, forces, and energy. **Particle physics** deals with the substructures of atoms and nuclei and is particularly aimed at finding those truly fundamental particles that have no further substructure. Just as in atomic and nuclear physics, we have found a complex array of particles and properties with systematic characteristics analogous to the periodic table and the chart of nuclides. An underlying structure is apparent, and there is some reason to think that we *are* finding particles that have no substructure. Of course, we have been in similar situations before. For example, atoms were once thought to be the ultimate substructures. It is possible that we could continue to find deeper and deeper structures without ever discovering the ultimate substructure—in science there is never complete certainty. See Figure 23.2.

The properties of matter are based on substructures called molecules and atoms. Each atom has the substructure of a nucleus surrounded by electrons, and their interactions explain atomic properties. Protons and neutrons—and the interactions between them—explain the stability and abundance of elements and form the substructure of nuclei. Protons and neutrons are not fundamental—they are composed of quarks. Like electrons and a few other particles, quarks may be the fundamental building blocks of all matter, lacking any further substructure. But the story is not complete because quarks and electrons may have substructures smaller than details that are presently observable.

Figure 23.2 A solid, a molecule, an atom, a nucleus, a nucleon (a particle that makes up the nucleus—either a proton or a neutron), and a quark.

This chapter covers the basics of particle physics as we know it today. An amazing convergence of topics is evolving in particle physics. We find that some particles are intimately related to forces and that nature on the smallest scale may have its greatest influence on the large scale character of the universe. It is an adventure exceeding the best science fiction because it is not only fantastic but also real.

23.1 The Four Fundamental Forces

Section Learning Objectives

By the end of the section, you will be able to do the following:

- Define, describe, and differentiate the four fundamental forces
- Describe the carrier particles and explain how their exchange transmits force
- Explain how particle accelerators work to gather evidence about particle physics

Section Key Terms

carrier particle	colliding beam	cyclotron	Feynman diagram	graviton
particle physics	pion	quantum electrodynamics	synchrotron	W^- boson
W^+ boson	weak nuclear force	Z^0 boson		

Despite the apparent complexity within the universe, there remain just four basic forces. These forces are responsible for all interactions known to science: from the very small to the very large to those that we experience in our day-to-day lives. These forces describe the movement of galaxies, the chemical reactions in our laboratories, the structure within atomic nuclei, and the cause of radioactive decay. They describe the true cause behind familiar terms like friction and the normal force. These four basic forces are known as fundamental because they alone are responsible for all observations of forces in nature. The four fundamental forces are gravity, electromagnetism, **weak nuclear force**, and strong nuclear force.

Understanding the Four Forces

The gravitational force is most familiar to us because it describes so many of our common observations. It explains why a dropped ball falls to the ground and why our planet orbits the Sun. It gives us the property of weight and determines much about the motion of objects in our daily lives. Because gravitational force acts between all objects of mass and has the ability to act over large distances, the gravitational force can be used to explain much of what we observe and can even describe the motion of objects on astronomical scales! That said, gravity is incredibly weak compared to the other fundamental forces and is the weakest of all of the fundamental forces. Consider this: The entire mass of Earth is needed to hold an iron nail to the ground. Yet with a simple magnet, the force of gravity can be overcome, allowing the nail to accelerate upward through space.

The electromagnetic force is responsible for both electrostatic interactions and the magnetic force seen between bar magnets. When focusing on the electrostatic relationship between two charged particles, the electromagnetic force is known as the coulomb force. The electromagnetic force is an important force in the chemical and biological sciences, as it is responsible for molecular connections like ionic bonding and hydrogen bonding. Additionally, the electromagnetic force is behind the common physics forces of friction and the normal force. Like the gravitational force, the electromagnetic force is an inverse square law. However, the electromagnetic force does not exist between any two objects of mass, only those that are charged.

When considering the structure of an atom, the electromagnetic force is somewhat apparent. After all, the electrons are held in place by an attractive force from the nucleus. But what causes the nucleus to remain intact? After all, if all protons are positive, it

makes sense that the coulomb force between the protons would repel the nucleus apart immediately. Scientists theorized that another force must exist within the nucleus to keep it together. They further theorized that this nuclear force must be significantly stronger than gravity, which has been observed and measured for centuries, and also stronger than the electromagnetic force, which would cause the protons to want to accelerate away from each other.

The strong nuclear force is an attractive force that exists between all nucleons. This force, which acts equally between proton-proton connections, proton-neutron connections, and neutron-neutron connections, is the strongest of all forces at short ranges. However, at a distance of 10^{-13} cm, or the diameter of a single proton, the force dissipates to zero. If the nucleus is large (it has many nucleons), then the distance between each nucleon could be much larger than the diameter of a single proton.

The weak nuclear force is responsible for beta decay, as seen in the equation ${}^{A}_{Z}X_N \rightarrow {}^{A}_{Z+1}Y_{N-1} + e + \nu$. Recall that beta decay is when a beta particle is ejected from an atom. In order to accelerate away from the nucleus, the particle must be acted on by a force. Enrico Fermi was the first to envision this type of force. While this force is appropriately labeled, it remains stronger than the gravitational force. However, its range is even smaller than that of the strong force, as can be seen in Table 23.1. The weak nuclear force is more important than it may appear at this time, as will be addressed when we discuss quarks.

Force	Approximate Relative Strength[1]	Range
Gravity	10^{-38}	∞
Weak	10^{-13}	$< 10^{-18}\,\text{m}$
Electromagnetic	10^{-2}	∞
Strong	1	$< 10^{-15}\,\text{m}$

[1]Relative strength is based on the strong force felt by a proton–proton pair.

Table 23.1 Relative strength and range of the four fundamental forces

Transmitting the Four Fundamental Forces

Just as it troubled Einstein prior to formulating the gravitational field theory, the concept of forces acting over a distance had greatly troubled particle physicists. That is, how does one proton *know* that another exists? Furthermore, what causes one proton to make a second proton repel? Or, for that matter, what is it about a proton that causes a neutron to attract? These mysterious interactions were first considered by Hideki Yukawa in 1935 and laid the foundation for much of what we now understand about particle physics.

Hideki Yukawa's focus was on the strong nuclear force and, in particular, its incredibly short range. His idea was a blend of particles, relativity, and quantum mechanics that was applicable to all four forces. Yukawa proposed that the nuclear force is actually transmitted by the exchange of particles, called **carrier particles**, and that what we commonly refer to as the force's field consists of these carrier particles. Specifically for the strong nuclear force, Yukawa proposed that a previously unknown particle, called a **pion**, is exchanged between nucleons, transmitting the force between them. Figure 23.3 illustrates how a pion would carry a force between a proton and a neutron.

Figure 23.3 The strong nuclear force is transmitted between a proton and neutron by the creation and exchange of a pion. The pion, created through a temporary violation of conservation of mass-energy, travels from the proton to the neutron and is recaptured. It is not

directly observable and is called a virtual particle. Note that the proton and neutron change identity in the process. The range of the force is limited by the fact that the pion can exist for only the short time allowed by the Heisenberg uncertainty principle. Yukawa used the finite range of the strong nuclear force to estimate the mass of the pion; the shorter the range, the larger the mass of the carrier particle.

In Yukawa's strong force, the carrier particle is assumed to be transmitted at the speed of light and is continually transferred between the two nucleons shown. The particle that Yukawa predicted was finally discovered within cosmic rays in 1947. Its name, the pion, stands for pi meson, where meson means *medium mass*; it's a medium mass because it is smaller than a nucleon but larger than an electron. Yukawa launched the field that is now called quantum chromodynamics, and the carrier particles are now called gluons due to their strong binding power. The reason for the change in the particle name will be explained when quarks are discussed later in this section.

As you may assume, the strong force is not the only force with a carrier particle. Nuclear decay from the weak force also requires a particle transfer. In the weak force are the following three: the weak negative carrier, W^-; the weak positive carrier, W^+; and the zero charge carrier, Z^0. As we will see, Fermi inferred that these particles must carry mass, as the total mass of the products of nuclear decay is slightly larger than the total mass of all reactants after nuclear decay.

The carrier particle for the electromagnetic force is, not surprisingly, the photon. After all, just as a lightbulb can emit photons from a charged tungsten filament, the photon can be used to transfer information from one electrically charged particle to another. Finally, the **graviton** is the proposed carrier particle for gravity. While it has not yet been found, scientists are currently looking for evidence of its existence (see Boundless Physics: Searching for the Graviton).

So how does a carrier particle transmit a fundamental force? Figure 23.4 shows a virtual photon transmitted from one positively charged particle to another. The transmitted photon is referred to as a virtual particle because it cannot be directly observed while transmitting the force. Figure 23.5 shows a way of graphing the exchange of a virtual photon between the two positively charged particles. This graph of time versus position is called a **Feynman diagram**, after the brilliant American physicist Richard Feynman (1918–1988), who developed it.

Figure 23.4 The image in part (a) shows the exchange of a virtual photon transmitting the electromagnetic force between charges, just as virtual pion exchange carries the strong nuclear force between nucleons. The image in part (b) shows that the photon cannot be directly observed in its passage because this would disrupt it and alter the force. In this case, the photon does not reach the other charge.

The Feynman diagram should be read from the bottom up to show the movement of particles over time. In it, you can see that the left proton is propelled leftward from the photon emission, while the right proton feels an impulse to the right when the photon is received. In addition to the Feynman diagram, Richard Feynman was one of the theorists who developed the field of **quantum electrodynamics** (QED), which further describes electromagnetic interactions on the submicroscopic scale. For this work, he shared the 1965 Nobel Prize with Julian Schwinger and S.I. Tomonaga. A Feynman diagram explaining the strong force interaction hypothesized by Yukawa can be seen in Figure 23.6. Here, you can see the change in particle type due to the exchange of the pi meson.

Figure 23.5 The Feynman diagram for the exchange of a virtual photon between two positively charged particles illustrates how electromagnetic force is transmitted on a quantum mechanical scale. Time is graphed vertically, while the distance is graphed horizontally. The two positively charged particles are seen to repel each other by the photon exchange.

Figure 23.6 The image shows a Feynman diagram for the exchange of a π+ (pion) between a proton and a neutron, carrying the strong nuclear force between them. This diagram represents the situation shown more pictorially in Figure 23.3.

The relative masses of the listed carrier particles describe something valuable about the four fundamental forces, as can be seen in Table 23.2. **W bosons** (consisting of W^- and W^+ bosons) and **Z bosons** (Z^0 bosons), carriers of the weak nuclear force, are nearly 1,000 times more massive than pions, carriers of the strong nuclear force. Simultaneously, the distance that the weak nuclear force can be transmitted is approximately $\frac{1}{1{,}000}$ times the strong force transmission distance. Unlike carrier particles, which have a limited range, the photon is a massless particle that has no limit to the transmission distance of the electromagnetic force. This relationship leads scientists to understand that the yet-unfound graviton is likely massless as well.

Force	Carrier Particle	Range	Relative Strength[1]
Gravity	Graviton (theorized)	∞	10^{-38}
Weak	W and Z bosons	∞	10^{-2}
Electromagnetic	Photon	$< 10^{-18}\,\text{m}$	10^{-13}
Strong	Pi mesons or pions (now known as gluons)	$< 10^{-15}\,\text{m}$	1

[1]Relative strength is based on the strong force felt by a proton-proton pair.

Table 23.2 Carrier particles and their relative masses compared to pions for the four fundamental forces

BOUNDLESS PHYSICS

Searching for the Graviton

From Newton's Universal Law of Gravitation to Einstein's field equations, gravitation has held the focus of scientists for centuries. Given the discovery of carrier particles during the twentieth century, the importance of understanding gravitation has yet again gained the interest of prominent physicists everywhere.

With carrier particles discovered for three of the four fundamental forces, it is sensible to scientists that a similar particle, titled the **graviton**, must exist for the gravitational force. While evidence of this particle is yet to be uncovered, scientists are working diligently to discover its existence.

So what do scientists think about the unfound particle? For starters, the graviton (like the photon) should be a massless particle traveling at the speed of light. This is assumed because, like the electromagnetic force, gravity is an inverse square law, $F \approx \frac{1}{r^2}$. Scientists also theorize that the graviton is an electrically neutral particle, as an empty space within the influence of gravity is chargeless.

However, because gravity is such a weak force, searching for the graviton has resulted in some unique methods. LIGO, the Laser Interferometer Gravitational-Wave Observatory, is one tool currently being utilized (see Figure 23.7). While searching for a gravitational wave to find a carrier particle may seem counterintuitive, it is similar to the approach taken by Planck and Einstein to learn more about the photon. According to wave-particle duality, if a gravitational wave can be found, the graviton should be present along with it. Predicted by Einstein's theory of general relativity, scientists have been monitoring binary star systems for evidence of these gravitational waves.

Figure 23.7 In searching for gravitational waves, scientists are using the Laser Interferometer Gravitational-Wave Observatory (LIGO). Here we see the control room of LIGO in Hanford, Washington.

Particle accelerators like the Large Hadron Collider (LHC) are being used to search for the graviton through high-energy collisions. While scientists at the LHC speculate that the particle may not exist long enough to be seen, evidence of its prior existence, like footprints in the sand, can be found through gaps in projected energy and momentum.

Some scientists are even searching the remnants of the Big Bang in an attempt to find the graviton. By observing the cosmic background radiation, they are looking for anomalies in gravitational waves that would provide information about the gravity particles that existed at the start of our universe.

Regardless of the method used, scientists should know the graviton once they find it. A massless, chargeless particle with a spin of 2 and traveling at the speed of light—there is no other particle like it. Should it be found, its discovery would surely be considered by future generations to be on par with those of Newton and Einstein.

GRASP CHECK

Why are binary star systems used by LIGO to find gravitational waves?

a. Binary star systems have high temperature.
b. Binary star systems have low density.
c. Binary star systems contain a large amount of mass, but because they are orbiting each other, the gravitational field between the two is much less.
d. Binary star systems contain a large amount of mass. As a result, the gravitational field between the two is great.

Accelerators Create Matter From Energy

Before looking at all the particles that make up our universe, let us first examine some of the machines that create them. The fundamental process in creating unknown particles is to accelerate known particles, such as protons or electrons, and direct a beam of them toward a target. Collisions with target nuclei provide a wealth of information, such as information obtained by Rutherford in the gold foil experiment. If the energy of the incoming particles is large enough, new matter can even be created in the collision. The more energy input or ΔE, the more matter m can be created, according to mass energy equivalence $m = \Delta E/c^2$. Limitations are placed on what can occur by known conservation laws, such as conservation of mass-energy, momentum, and charge. Even more interesting are the unknown limitations provided by nature. While some expected reactions do occur, others do not, and still other unexpected reactions may appear. New laws are revealed, and the vast majority of what we know about particle physics has come from accelerator laboratories. It is the particle physicist's favorite indoor sport.

Our earliest model of a particle accelerator comes from the Van de Graaff generator. The relatively simple device, which you have likely seen in physics demonstrations, can be manipulated to produce potentials as great as 50 million volts. While these machines do not have energies large enough to produce new particles, analysis of their accelerated ions was instrumental in exploring several aspects of the nucleus.

Another equally famous early accelerator is the **cyclotron**, invented in 1930 by the American physicist, E.O. Lawrence (1901–1958). Figure 23.8 is a visual representation with more detail. Cyclotrons use fixed-frequency alternating electric fields to accelerate particles. The particles spiral outward in a magnetic field, making increasingly larger radius orbits during acceleration. This clever arrangement allows the successive addition of electric potential energy with each loop. As a result, greater particle energies are possible than in a Van de Graaff generator.

Figure 23.8 On the left is an artist's rendition of the popular physics demonstration tool, the Van de Graaff generator. A battery (A) supplies excess positive charge to a pointed conductor, the points of which spray the charge onto a moving insulating belt near the bottom. The pointed conductor (B) on top in the large sphere picks up the charge. (The induced electric field at the points is so large that it removes the charge from the belt.) This can be done because the charge does not remain inside the conducting sphere but moves to its outer surface. An ion source inside the sphere produces positive ions, which are accelerated away from the positive sphere to high velocities. On the right is a cyclotron. Cyclotrons use a magnetic field to cause particles to move in circular orbits. As the particles pass between the plates of the Dees, the voltage across the gap is oscillated to accelerate them twice in each orbit.

A **synchrotron** is a modification of the cyclotron in which particles continually travel in a fixed-radius orbit, increasing speed each time. Accelerating voltages are synchronized with the particles to accelerate them, hence the name. Additionally, magnetic field strength is increased to keep the orbital radius constant as energy increases. A ring of magnets and accelerating tubes, as shown in Figure 23.9, are the major components of synchrotrons. High-energy particles require strong magnetic fields to steer

them, so superconducting magnets are commonly employed. Still limited by achievable magnetic field strengths, synchrotrons need to be very large at very high energies since the radius of a high-energy particle's orbit is very large.

To further probe the nucleus, physicists need accelerators of greater energy and detectors of shorter wavelength. To do so requires not only greater funding but greater ingenuity as well. **Colliding beams** used at both the Fermi National Accelerator Laboratory (Fermilab; see Figure 23.11) near Chicago and the LHC in Switzerland are designed to reduce energy loss in particle collisions. Typical stationary particle detectors lose a large amount of energy to the recoiling target struck by the accelerating particle. By providing head-on collisions between particles moving in opposite directions, colliding beams make it possible to create particles with momenta and kinetic energies near zero. This allows for particles of greater energy and mass to be created. Figure 23.10 is a schematic representation of this effect. In addition to circular accelerators, linear accelerators can be used to reduce energy radiation losses. The Stanford Linear Accelerator Center (now called the SLAC National Accelerator Laboratory) in California is home to the largest such accelerator in the world.

Figure 23.9 (a) A synchrotron has a ring of magnets and accelerating tubes. The frequency of the accelerating voltages is increased to cause the beam particles to travel the same distance in a shorter time. The magnetic field should also be increased to keep each beam burst traveling in a fixed-radius path. Limits on magnetic field strength require these machines to be very large in order to accelerate particles to very high energies. (b) A positively charged particle is shown in the gap between accelerating tubes. (c) While the particle passes through the tube, the potentials are reversed so that there is another acceleration at the next gap. The frequency of the reversals needs to be varied as the particle is accelerated to achieve successive accelerations in each gap.

Figure 23.10 This schematic shows the two rings of Fermilab's accelerator and the scheme for colliding protons and antiprotons (not to scale).

Figure 23.11 The Fermi National Accelerator Laboratory, near Batavia, Illinois, was a subatomic particle collider that accelerated protons and antiprotons to attain energies up to 1 Tev (a trillion electronvolts). The circular ponds near the rings were built to dissipate waste heat. This accelerator was shut down in September 2011. (credit: Fermilab, Reidar Hahn)

Check Your Understanding

1. Which of the four forces is responsible for radioactive decay?
   a. the electromagnetic force

b. the gravitational force
c. the strong nuclear force
d. the weak nuclear force

2. What force or forces exist between an electron and a proton?
   a. the strong nuclear force, the electromagnetic force, and gravity
   b. the weak nuclear force, the strong nuclear force, and gravity
   c. the weak nuclear force, the strong nuclear force, and the electromagnetic force
   d. the weak nuclear force, the electromagnetic force, and gravity

3. What is the proposed carrier particle for the gravitational force?
   a. boson
   b. graviton
   c. gluon
   d. photon

4. What is the relationship between the mass and range of a carrier particle?
   a. Range of a carrier particle is inversely proportional to its mass.
   b. Range of a carrier particle is inversely proportional to square of its mass.
   c. Range of a carrier particle is directly proportional to its mass.
   d. Range of a carrier particle is directly proportional to square of its mass.

5. What type of particle accelerator uses fixed-frequency oscillating electric fields to accelerate particles?
   a. cyclotron
   b. synchrotron
   c. betatron
   d. Van de Graaff accelerator

6. How does the expanding radius of the cyclotron provide evidence of particle acceleration?
   a. A constant magnetic force is exerted on particles at all radii. As the radius increases, the velocity of the particle must increase to maintain this constant force.
   b. A constant centripetal force is exerted on particles at all radii. As the radius increases, the velocity of the particle must decrease to maintain this constant force.
   c. A constant magnetic force is exerted on particles at all radii. As the radius increases, the velocity of the particle must decrease to maintain this constant force.
   d. A constant centripetal force is exerted on particles at all radii. As the radius increases, the velocity of the particle must increase to maintain this constant force.

7. Which of the four forces is responsible for the structure of galaxies?
   a. electromagnetic force
   b. gravity
   c. strong nuclear force
   d. weak nuclear force

23.2 Quarks

Section Learning Objectives

By the end of the section, you will be able to do the following:

- Describe quarks and their relationship to other particles
- Distinguish hadrons from leptons
- Distinguish matter from antimatter
- Describe the standard model of the atom
- Define a Higgs boson and its importance to particle physics

Section Key Terms

annihilation	antimatter	baryon	bottom quark	charmed quark
color	down quark	flavor	gluon	hadron
Higgs boson	Higgs field	lepton	meson	pair production
positron	quantum chromodynamics	quark	Standard Model	strange quark
top quark	up quark			

Quarks

"The first principles of the universe are atoms and empty space. Everything else is merely thought to exist…"

"… Further, the atoms are unlimited in size and number, and they are borne along with the whole universe in a vortex, and thereby generate all composite things—fire, water, air, earth. For even these are conglomerations of given atoms. And it because of their solidity that these atoms are impassive and unalterable."

—Diogenes Laertius (summarizing the views of Democritus, circa 460–370 B.C.)

The search for fundamental particles is nothing new. Atomists of the Greek and Indian empires, like Democritus of fifth century B.C., openly wondered about the most finite components of our universe. Though dormant for centuries, curiosity about the atomic nature of matter was reinvigorated by Rutherford's gold foil experiment and the discovery of the nucleus. By the early 1930s, scientists believed they had fully determined the tiniest constituents of matter—in the form of the proton, neutron, and electron.

This would be only partially true. At present, scientists know that there are hundreds of particles not unlike our electron and nucleons, all making up what some have termed the *particle zoo*. While we are confident that the electron remains fundamental, it is surrounded by a plethora of similar sounding terms, like leptons, hadrons, baryons, and mesons. Even though not every particle is considered fundamental, they all play a vital role in understanding the intricate structure of our universe.

A fundamental particle is defined as a particle with no substructure and no finite size. According to the **Standard Model**, there are three types of fundamental particles: leptons, quarks, and carrier particles. As you may recall, carrier particles are responsible for transmitting fundamental forces between their interacting masses. **Leptons** are a group of six particles not bound by the strong nuclear force, of which the electron is one. As for quarks, they are the fundamental building blocks of a group of particles called **hadrons**, a group that includes both the proton and the neutron.

Now for a brief history of **quarks**. Quarks were first proposed independently by American physicists Murray Gell-Mann and George Zweig in 1963. Originally, three quark types—or **flavors**—were proposed with the names **up** (u), **down** (d), and **strange** (s).

At first, physicists expected that, with sufficient energy, we should be able to free quarks and observe them directly. However, this has not proved possible, as the current understanding is that the force holding quarks together is incredibly great and, much like a spring, increases in magnitude as the quarks are separated. As a result, when large energies are put into collisions, other particles are created—but no quarks emerge. With that in mind, there is compelling evidence for the existence of quarks. By 1967, experiments at the SLAC National Accelerator Laboratory scattering 20-GeV electrons from protons produced results like Rutherford had obtained for the nucleus nearly 60 years earlier. The SLAC scattering experiments showed unambiguously that there were three point-like (meaning they had sizes considerably smaller than the probe's wavelength) charges inside the proton as seen in Figure 23.12. This evidence made all but the most skeptical admit that there was validity to the quark substructure of hadrons.

Figure 23.12 Scattering of high-energy electrons from protons at facilities like SLAC produces evidence of three point-like charges consistent with proposed quark properties. This experiment is analogous to Rutherford's discovery of the small size of the nucleus by scattering α particles. High-energy electrons are used so that the probe wavelength is small enough to see details smaller than the proton.

The inclusion of the strange quark with Zweig and Gell-Mann's model concerned physicists. While the up and down quarks demonstrated fairly clear symmetry and were present in common fundamental particles like protons and neutrons, the strange quark did not have a counterpart of its own. This thought, coupled with the four known leptons at the time, caused scientists to predict that a fourth quark, yet to be found, also existed.

In 1974, two groups of physicists independently discovered a particle with this new quark, labeled **charmed**. This completed the second *exotic* quark pair, strange (s) and charmed (c). A final pair of quarks was proposed when a third pair of leptons was discovered in 1975. The existence of the **bottom** (b) quark and the **top** (t) quark was verified through experimentation in 1976 and 1995, respectively. While it may seem odd that so much time would elapse between the original quark discovery in 1967 and the verification of the top quark in 1995, keep in mind that each quark discovered had a progressively larger mass. As a result, each new quark has required more energy to discover.

TIPS FOR SUCCESS

Note that a very important tenet of science occurred throughout the period of quark discovery. The charmed, bottom, and top quarks were all speculated on, and then were discovered some time later. Each of their discoveries helped to verify and strengthen the quark model. This process of speculation and verification continues to take place today and is part of what drives physicists to search for evidence of the graviton and Grand Unified Theory.

One of the most confounding traits of quarks is their electric charge. Long assumed to be discrete, and specifically a multiple of the elementary charge of the electron, the electric charge of an individual quark is fractional and thus seems to violate a presumed tenet of particle physics. The fractional charge of quarks, which are $\pm\left(\frac{2}{3}\right)q_e$ and $\pm\left(\frac{1}{3}\right)q_e$, are the only structures found in nature with a nonintegral number of charge q. However, note that despite this odd construction, the fractional value of the quark does not violate the quantum nature of the charge. After all, free quarks cannot be found in nature, and all quarks are bound into arrangements in which an integer number of charge is constructed. Table 23.3 shows the six known quarks, in addition to their antiquark components, as will be discussed later in this section.

Flavor	Symbol	Antiparticle	Charge[1][2]
Up	u	$\overline{u}$	$\pm\frac{2}{3}q_e$
Down	d	$\overline{d}$	$\mp\frac{1}{3}q_e$
Strange	s	$\overline{s}$	$\mp\frac{1}{3}q_e$
Charmed	c	$\overline{c}$	$\pm\frac{2}{3}q_e$

[1]The lower of the $\pm$ symbols are the values for antiquarks.
[2]There are further qualities that differentiate between quarks. However, they are beyond the discussion in this text.

Table 23.3 Quarks and Antiquarks

Flavor	Symbol	Antiparticle	Charge[1][2]
Bottom	b	$\overline{b}$	$\mp\frac{1}{3}q_e$
Top	t	$\overline{t}$	$\pm\frac{2}{3}q_e$

[1]The lower of the $\pm$ symbols are the values for antiquarks.
[2]There are further qualities that differentiate between quarks. However, they are beyond the discussion in this text.

Table 23.3 Quarks and Antiquarks

While the term *flavor* is used to differentiate between types of quarks, the concept of **color** is more analogous to the electric charge in that it is primarily responsible for the force interactions between quarks. Note—Take a moment to think about the electrostatic force. It is the electric charge that causes attraction and repulsion. It is the same case here but with a *color* charge. The three colors available to a quark are red, green, and blue, with antiquarks having colors of anti-red (or cyan), anti-green (or magenta), and anti-blue (or yellow).

Why use colors when discussing quarks? After all, the quarks are not actually colored with visible light. The reason colors are used is because the properties of a quark are analogous to the three primary and secondary colors mentioned above. Just as different colors of light can be combined to create white, different *colors* of quark may be combined to construct a particle like a proton or neutron. In fact, for each hadron, the quarks must combine such that their color sums to white! Recall that two up quarks and one down quark construct a proton, as seen in Figure 23.12. The sum of the three quarks' colors—red, green, and blue—yields the color white. This theory of color interaction within particles is called **quantum chromodynamics**, or QCD. As part of QCD, the strong nuclear force can be explained using color. In fact, some scientists refer to the color force, not the strong force, as one of the four fundamental forces. Figure 23.13 is a Feynman diagram showing the interaction between two quarks by using the transmission of a colored **gluon**. Note that the gluon is also considered the charge carrier for the strong nuclear force.

Figure 23.13 The exchange of gluons between quarks carries the strong force and may change the color of the interacting quarks. While the colors of the individual quarks change, their flavors do not.

Note that quark flavor may have any color. For instance, in Figure 23.13, the down quark has a red color and a green color. In other words, colors are not specific to a particle quark flavor.

Hadrons and Leptons

Particles can be revealingly grouped according to what forces they feel between them. All particles (even those that are massless) are affected by gravity since gravity affects the space and time in which particles exist. All charged particles are affected by the electromagnetic force, as are neutral particles that have an internal distribution of charge (such as the neutron with its magnetic moment). Special names are given to particles that feel the strong and weak nuclear forces. Hadrons are particles that feel the strong nuclear force, whereas leptons are particles that do not. All particles feel the weak nuclear force. This means that hadrons are distinguished by being able to feel both the strong and weak nuclear forces. Leptons and hadrons are distinguished in other ways as well. Leptons are fundamental particles that have no measurable size, while hadrons are composed of quarks and have a diameter on the order of 10 to 15 m. Six particles, including the electron and neutrino, make up the list of known leptons. There are hundreds of complex particles in the hadron class, a few of which (including the proton and neutron) are listed in Table 23.4.

Category	Particle Name	Symbol	Antiparticle	Rest Mass $(\mathrm{MeV/c^2})$	Mean Lifetime (s)
Leptons	Electron	e^-	e^+	0.511	Stable
	Neutrino (e)	ν_e	$\bar{\nu}_e$	0 (7.0 eV) [1]	Stable
	Muon	μ^-	μ^+	105.7	2.20×10^{-6}
	Neutrino (μ)	ν_μ	$\bar{\nu}_\mu$	0 (<0.27) [1]	Stable
	Tau	τ^-	τ^+	1,777	2.91×10^{-6}
	Neutrino (τ)	ν_τ	$\bar{\nu}_\tau$	0 (<31) [1]	Stable
Hadrons – Mesons[2]	Pion	π^+	π^-	139.6	2.60×10^{-8}
		π^0	Self	135.0	8.40×10^{-17}
	Kaon	K^+	K^-	493.7	1.24×10^{-8}
		K^0	K^0	497.6	0.90×10^{-10}
	Eta	η^0	Self	547.9	2.53×10^{-19}
Hadrons – Baryons[3]	Proton	p	$\bar{p}$	938.3	Stable
	Neutron	n	$\bar{n}$	939.6	882
	Lambda	Λ^0	$\bar{\Lambda}^0$	1,115.7	2.63×10^{-10}
	Omega	Ω^-	Ω^+	1,672.5	0.82×10^{-10}

[1]Neutrino masses may be zero. Experimental upper limits are given in parentheses.
[2]Many other mesons known
[3]Many other baryons known

Table 23.4 List of Leptons and Hadrons.

There are many more leptons, mesons, and baryons yet to be discovered and measured. The purpose of trying to uncover the smallest indivisible things in existence is to explain the world around us through forces and the interactions between particles, galaxies and objects. This is why a handful of scientists devote their life's work to smashing together small particles.

What internal structure makes a proton so different from an electron? The proton, like all hadrons, is made up of quarks. A few examples of hadron quark composition can be seen in Figure 23.14. As shown, each hadron is constructed of multiple quarks. As mentioned previously, the fractional quark charge in all four hadrons sums to the particle's integral value. Also, notice that the color composition for each of the four particles adds to white. Each of the particles shown is constructed of up, down, and their antiquarks. This is not surprising, as the quarks strange, charmed, top, and bottom are found in only our most exotic particles.

Figure 23.14 All baryons, such as the proton and neutron shown here, are composed of three quarks. All mesons, such as the pions shown here, are composed of a quark–antiquark pair. Arrows represent the spins of the quarks. The colors are such that they need to add to white for any possible combination of quarks.

You may have noticed that while the proton and neutron in Figure 23.14 are composed of three quarks, both pions are comprised of only two quarks. This refers to a final delineation in particle structure. Particles with three quarks are called **baryons**. These are heavy particles that can decay into another baryon. Particles with only two quarks—a-quark–anti-quark pair—are called **mesons**. These are particles of moderate mass that cannot decay into the more massive baryons.

Before continuing, take a moment to view Figure 23.15. In this figure, you can see the strong force reimagined as a color force. The particles interacting in this figure are the proton and neutron, just as they were in Figure 23.6. This reenvisioning of the strong force as an interaction between colored quarks is the critical concept behind quantum chromodynamics.

Figure 23.15 This Feynman diagram shows the interaction between a proton and a neutron, corresponding to the interaction shown in Figure 23.6. This diagram, however, shows the quark and gluon details of the strong nuclear force interaction.

Matter and Antimatter

Antimatter was first discovered in the form of the **positron**, the positively charged electron. In 1932, American physicist Carl Anderson discovered the positron in cosmic ray studies. Through a cloud chamber modified to curve the trajectories of cosmic

rays, Anderson noticed that the curves of some particles followed that of a negative charge, while others curved like a positive charge. However, the positive curve showed not the mass of a proton but the mass of an electron. This outcome is shown in Figure 23.16 and suggests the existence of a positively charged version of the electron, created by the destruction of solar photons.

Figure 23.16 The image above is from the Fermilab *15 foot bubble chamber* and shows the production of an electron and positron (or antielectron) from an incident photon. This event is titled **pair production** and provides evidence of antimatter, as the two repel each other.

Antimatter is considered the opposite of matter. For most antiparticles, this means that they share the same properties as their original particles with the exception of their charge. This is why the positron can be considered a positive electron while the antiproton is considered a negative proton. The idea of an opposite charge for neutral particles (like the neutron) can be confusing, but it makes sense when considered from the quark perspective. Just as the neutron is composed of one up quark and two down quarks (of charge $+\frac{2}{3}$ and $-\frac{1}{3}$, respectively), the antineutron is composed of one anti–up quark and two anti–down quarks (of charge $-\frac{2}{3}$ and $+\frac{1}{3}$, respectively). While the overall charge of the neutron remains the same, its constituent particles do not!

A word about antiparticles: Like regular particles, antiparticles could function just fine on their own. In fact, a universe made up of antimatter may operate just as our own matter-based universe does. However, we do not know fully whether this is the case. The reason for this is **annihilation**. Annihilation is the process of destruction that occurs when a particle and its antiparticle interact. As soon as two particles (like a positron and an electron) coincide, they convert their masses to energy through the equation $E = mc^2$. This mass-to-energy conversion, which typically results in photon release, happens instantaneously and makes it very difficult for scientists to study antimatter. That said, scientists have had success creating antimatter through high-energy particle collisions. Both antineutrons and antiprotons were created through accelerator experiments in 1956, and an anti–hydrogen atom was even created at CERN in 1995! As referenced in , the annihilation of antiparticles is currently used in medical studies to determine the location of radioisotopes.

Completing the Standard Model of the Atom

The Standard Model of the atom refers to the current scientific view of the fundamental components and interacting forces of matter. The Standard Model (Figure 23.17) shows the six quarks that bind to form all hadrons, the six lepton particles already considered fundamental, the four carrier particles (or gauge bosons) that transmit forces between the leptons and quarks, and the recently added **Higgs boson** (which will be discussed shortly). This totals 17 fundamental particles, combinations of which are responsible for all known matter in our entire universe! When adding the antiquarks and antileptons, 31 components make up the Standard Model.

Figure 23.17 The Standard Model of elementary particles shows an organized view of all fundamental particles, as currently known: six quarks, six leptons, and four gauge bosons (or carrier particles). The Higgs boson, first observed in 2012, is a new addition to the Standard Model.

Figure 23.17 shows all particles within the Standard Model of the atom. Not only does this chart divide all known particles by color-coded group, but it also provides information on particle stability. Note that the color-coding system in this chart is separate from the red, green, and blue color labeling system of quarks. The first three columns represent the three *families* of matter. The first column, considered Family 1, represents particles that make up normal matter, constructing the protons, neutrons, and electrons that make up the common world. Family 2, represented from the charm quark to the muon neutrino, is comprised of particles that are more massive. The leptons in this group are less stable and more likely to decay. Family 3, represented by the third column, are more massive still and decay more quickly. The order of these families also conveniently represents the order in which these particles were discovered.

TIPS FOR SUCCESS

Look for trends that exist within the Standard Model. Compare the charge of each particle. Compare the spin. How does mass relate to the model structure? Recognizing each of these trends and asking questions will yield more insight into the organization of particles and the forces that dictate particle relationships. Our understanding of the Standard Model is still young, and the questions you may have in analyzing the Standard Model may be some of the same questions that particle physicists are searching for answers to today!

The Standard Model also summarizes the fundamental forces that exist as particles interact. A closer look at the Standard Model, as shown in Figure 23.18, reveals that the arrangement of carrier particles describes these interactions.

Figure 23.18 The revised Standard Model shows the interaction between gauge bosons and other fundamental particles. These interactions are responsible for the fundamental forces, three of which are described through the chart's shaded areas.

Each of the shaded areas represents a fundamental force and its constituent particles. The red shaded area shows all particles involved in the strong nuclear force, which we now know is due to quantum chromodynamics. The blue shaded area corresponds to the electromagnetic force, while the green shaded area corresponds to the weak nuclear force, which affects all quarks and leptons. The electromagnetic force and weak nuclear force are considered united by the electroweak force within the Standard Model. Also, because definitive evidence of the graviton is yet to be found, it is not included in the Standard Model.

The Higgs Boson

One interesting feature of the Standard Model shown in Figure 23.18 is that, while the gluon and photon have no mass, the Z and W bosons are very massive. What supplies these quickly moving particles with mass and not the gluons and photons? Furthermore, what causes some quarks to have more mass than others?

In the 1960s, British physicist Peter Higgs and others speculated that the W and Z bosons were actually just as massless as the gluon and photon. However, as the W and Z bosons traveled from one particle to another, they were slowed down by the presence of a **Higgs field**, much like a fish swimming through water. The thinking was that the existence of the Higgs field would slow down the bosons, causing them to decrease in energy and thereby transfer this energy to mass. Under this theory, all particles pass through the Higgs field, which exists throughout the universe. The gluon and photon travel through this field as well but are able to do so unaffected.

The presence of a force from the Higgs field suggests the existence of its own carrier particle, the Higgs boson. This theorized boson interacts with all particles but gluons and photons, transferring force from the Higgs field. Particles with large mass (like the top quark) are more likely to receive force from the Higgs boson.

While it is difficult to examine a field, it is somewhat simpler to find evidence of its carrier. On July 4, 2012, two groups of scientists at the LHC independently confirmed the existence of a Higgs-like particle. By examining trillions of proton–proton collisions at energies of 7 to 8 TeV, LHC scientists were able to determine the constituent particles that created the protons. In this data, scientists found a particle with similar mass, spin, parity, and interactions with other particles that matched the Higgs boson predicted decades prior. On March 13, 2013, the existence of the Higgs boson was tentatively confirmed by CERN. Peter Higgs and Francois Englert received the Nobel Prize in 2013 for the "theoretical discovery of a mechanism that contributes to our understanding of the origin and mass of subatomic particles."

WORK IN PHYSICS

Particle Physicist

If you have an innate desire to unravel life's great mysteries and further understand the nature of the physical world, a career in particle physics may be for you!

Particle physicists have played a critical role in much of society's technological progress. From lasers to computers, televisions to space missions, splitting the atom to understanding the DNA molecule to MRIs and PET scans, much of our modern society is based on the work done by particle physicists.

While many particle physicists focus on specialized tasks in the fields of astronomy and medicine, the main goal of particle physics is to further scientists' understanding of the Standard Model. This may mean work in government, industry, or

academics. Within the government, jobs in particle physics can be found within the National Institute for Standards and Technology, Department of Energy, NASA, and Department of Defense. Both the electronics and computer industries rely on the expertise of particle physicists. College teaching and research positions can also be potential career opportunities for particle physicists, though they often require some postgraduate work as a prerequisite. In addition, many particle physicists are employed to work on high-energy colliders. Domestic collider labs include the Brookhaven National Laboratory in New York, the Fermi National Accelerator Laboratory near Chicago, and the SLAC National Accelerator Laboratory operated by Stanford University. For those who like to travel, work at international collider labs can be found at the CERN facility in Switzerland in addition to institutes like the Budker Institute of Nuclear Physics in Russia, DESY in Germany, and KEK in Japan.

Shirley Jackson became the first African American woman to earn a Ph.D. from MIT back in 1973, and she went on to lead a highly successful career in the field of particle physics. Like Dr. Jackson, successful students of particle physics grow up with a strong curiosity in the world around them and a drive to continually learn more. If you are interested in exploring a career in particle physics, work to achieve good grades and SAT scores, and find time to read popular books on physics topics that interest you. While some math may be challenging, recognize that this is only a tool of physics and should not be considered prohibitive to the field. High-level work in particle physics often requires a Ph.D.; however, it is possible to find work with a master's degree. Additionally, jobs in industry and teaching can be achieved with solely an undergraduate degree.

GRASP CHECK

What is the primary goal of all work in particle physics?

a. The primary goal is to further our understanding of the Standard Model.
b. The primary goal is to further our understanding of Rutherford's model.
c. The primary goal is to further our understanding of Bohr's model.
d. The primary goal is to further our understanding of Thomson's model.

Check Your Understanding

8. In what particle were quarks originally discovered?

a. the electron
b. the neutron
c. the proton
d. the photon

9. Why was the existence of the charm quark speculated, even though no direct evidence of it existed?

a. The existence of the charm quark was symmetrical with up and down quarks. Additionally, there were two known leptons at the time and only two quarks.
b. The strange particle lacked the symmetry that existed with the up and down quarks. Additionally, there were four known leptons at the time and only three quarks.
c. The bottom particle lacked the symmetry that existed with the up and down quarks. Additionally, there were two known leptons at the time and only two quarks.
d. The existence of charm quarks was symmetrical with up and down quarks. Additionally, there were four known leptons at the time and only three quarks.

10. What type of particle is the electron?

a. The electron is a lepton.
b. The electron is a hadron.
c. The electron is a baryon.
d. The electron is an antibaryon.

11. How do the number of fundamental particles differ between hadrons and leptons?

a. Hadrons are constructed of at least three fundamental quark particles, while leptons are fundamental particles.
b. Hadrons are constructed of at least three fundamental quark particles, while leptons are constructed of two fundamental particles.
c. Hadrons are constructed of at least two fundamental quark particles, while leptons are constructed of three

fundamental particles.

d. Hadrons are constructed of at least two fundamental quark particles, while leptons are fundamental particles.

12. Does antimatter exist?

a. no
b. yes

13. How does the deconstruction of a photon into an electron and a positron uphold the principles of mass and charge conservation?

a. The sum of the masses of an electron and a positron is equal to the mass of the photon before pair production. The sum of the charges on an electron and a positron is equal to the zero charge of the photon.
b. The sum of the masses of an electron and a positron is equal to the mass of the photon before pair production. The sum of the same charges on an electron and a positron is equal to the charge on a photon.
c. During the particle production the total energy of the photon is converted to the mass of an electron and a positron. The sum of the opposite charges on the electron and positron is equal to the zero charge of the photon.
d. During particle production, the total energy of the photon is converted to the mass of an electron and a positron. The sum of the same charges on an electron and a positron is equal to the charge on a photon.

14. How many fundamental particles exist in the Standard Model, including the Higgs boson and the graviton (not yet observed)?

a. 12
b. 15
c. 13
d. 19

15. Why do gluons interact only with particles in the first two rows of the Standard Model?

a. The leptons in the third and fourth rows do not have mass, but the gluons can interact between the quarks through gravity only.
b. The leptons in the third and fourth rows do not have color, but the gluons can interact between quarks through color interactions only.
c. The leptons in the third and fourth rows do not have spin, but the gluons can interact between quarks through spin interactions only.
d. The leptons in the third and fourth rows do not have charge, but the gluons can interact between quarks through charge interactions only.

16. What fundamental property is provided by particle interaction with the Higgs boson?

a. charge
b. mass
c. spin
d. color

17. Considering the Higgs field, what differentiates more massive particles from less massive particles?

a. More massive particles interact more with the Higgs field than the less massive particles.
b. More massive particles interact less with the Higgs field than the less massive particles.

18. What particles were launched into the proton during the original discovery of the quark?

a. bosons
b. electrons
c. neutrons
d. photons

23.3 The Unification of Forces

Section Learning Objectives

By the end of the section, you will be able to do the following:

- Define a grand unified theory and its importance
- Explain the evolution of the four fundamental forces from the Big Bang onward
- Explain how grand unification theories can be tested

Section Key Terms

Big Bang	Electroweak Epoch	electroweak theory	Grand Unification Epoch	Grand Unified Theory
Inflationary Epoch	Planck Epoch	Quark Era	superforce	Theory of Everything

Understanding the Grand Unified Theory

Present quests to show that the four basic forces are different manifestations of a single unified force that follow a long tradition. In the nineteenth century, the distinct electric and magnetic forces were shown to be intimately connected and are now collectively called the electromagnetic force. More recently, the weak nuclear force was united with the electromagnetic force. As shown in Figure 23.19, carrier particles transmit three of the four fundamental forces in very similar ways. With these considerations in mind, it is natural to suggest that a theory may be constructed in which the strong nuclear, weak nuclear, and electromagnetic forces are all unified. The search for a correct theory linking the forces, called the **Grand Unified Theory (GUT)**, is explored in this section.

In the 1960s, the **electroweak theory** was developed by Steven Weinberg, Sheldon Glashow, and Abdus Salam. This theory proposed that the electromagnetic and weak nuclear forces are identical at sufficiently high energies. At lower energies, like those in our present-day universe, the two forces remain united but manifest themselves in different ways. One of the main consequences of the electroweak theory was the prediction of three short-range carrier particles, now known as the W^+, W^-, and Z^0 bosons. Not only were three particles predicted, but the mass of each W^+ and W^- boson was predicted to be 81 GeV/c^2, and that of the Z^0 boson was predicted to be 90 GeV/c^2. In 1983, these carrier particles were observed at CERN with the predicted characteristics, including masses having those predicted values as given in .

How can forces be unified? They are definitely distinct under most circumstances. For example, they are carried by different particles and have greatly different strengths. But experiments show that at extremely short distances and at extremely high energies, the strengths of the forces begin to become more similar, as seen in Figure 23.20.

Figure 23.19 The exchange of a virtual Z^0 particle (boson) carries the weak nuclear force between an electron and a neutrino in this Feynman diagram. This diagram is similar to the diagrams in Figure 23.6 and for the electromagnetic and strong nuclear forces.

As discussed earlier, the short ranges and large masses of the weak carrier bosons require correspondingly high energies to create them. Thus, the energy scale on the horizontal axis of Figure 23.20 also corresponds to shorter and shorter distances

(going from left to right), with 100 GeV corresponding to approximately 10^{-18} m, for example. At that distance, the strengths of the electromagnetic and weak nuclear forces are the same. To test this, energies of about 100 GeV are put into the system. When this occurs, the W^+, W^-, and Z^0 carrier particles are created and released. At those and higher energies, the masses of the carrier particles become less and less relevant, and the Z^0 boson in particular resembles the massless, chargeless photon. As further energy is added, the W^+, W^-, and Z^0 particles are further transformed into massless carrier particles even more similar to photons and gluons.

Figure 23.20 The relative strengths of the four basic forces vary with distance, and, hence, energy is needed to probe small distances. At ordinary energies (a few eV or less), the forces differ greatly. However, at energies available in accelerators, the weak nuclear and electromagnetic (EM) forces become unified. Unfortunately, the energies at which the strong nuclear and electroweak forces become the same are unreachable in any conceivable accelerator. The universe may provide a laboratory, and nature may show effects at ordinary energies that give us clues about the validity of this graph.

The extremely short distances and high energies at which the electroweak force becomes identical with the strong nuclear force are not reachable with any conceivable human-built accelerator. At energies of about 10^{14} GeV (16,000 J per particle), distances of about 10 to 30 m can be probed. Such energies are needed to test the theory directly, but these are about 10^{10} times higher than the maximum energy associated with the LHC, and the distances are about 10 to 12 smaller than any structure we have direct knowledge of. This would be the realm of various GUTs, of which there are many, since there is no constraining evidence at these energies and distances. Past experience has shown that anytime you probe so many orders of magnitude further, you find the unexpected.

While direct evidence of a GUT is not presently possible, that does not rule out the ability to assess a GUT through an indirect process. Current GUTs require various other events as a consequence of their theory. Some GUTs require the existence of magnetic monopoles, very massive individual north- and south-pole particles, which have not yet been proven to exist, while others require the use of extra dimensions. However, not all theories result in the same consequences. For example, disproving the existence of magnetic monopoles will not disprove all GUTs. Much of the science we accept in our everyday lives is based on different models, each with their own strengths and limitations. Although a particular model may have drawbacks, that does not necessarily mean that it should be discounted completely.

One consequence of GUTs that can theoretically be assessed is proton decay. Multiple current GUTs hypothesize that the stable proton should actually decay at a lifetime of 10^{31} years. While this time is incredibly large (keep in mind that the age of the universe is less than 14 billion years), scientists at the Super-Kamiokande in Japan have used a 50,000-ton tank of water to search for its existence. The decay of a single proton in the Super-Kamiokande tank would be observed by a detector, thereby providing support for the predicting GUT model. However, as of 2014, 17 years into the experiment, decay is yet to be found. This time span equates to a minimum limit on proton life of 5.9×10^{33} years. While this result certainly does not support many grand unifying theories, an acceptable model may still exist.

TIPS FOR SUCCESS

The Super-Kamiokande experiment is a clever use of proportional reasoning. Because it is not feasible to test for 10^{31} years in order for a single proton to decay, scientists chose instead to manipulate the proton–time ratio. If one proton decays in 10^{31}

years, then in one year 10^{-31} protons will decay. With this in mind, if scientists wanted to test the proton decay theory in one year, they would need 10^{31} protons. While this is also unfeasible, the use of a 50,000-ton tank of water helps to bring both the wait time and proton number to within reason.

The Standard Model and the Big Bang

Nature is full of examples where the macroscopic and microscopic worlds intertwine. Newton realized that the nature of gravity on Earth that pulls an apple to the ground could explain the motion of the moon and planets so much farther away. Decays of tiny nuclei explain the hot interior of the Earth. Fusion of nuclei likewise explains the energy of stars. Today, the patterns in particle physics seem to be explaining the evolution and character of the universe. And the nature of the universe has implications for unexplored regions of particle physics.

In 1929, Edwin Hubble observed that all but the closest galaxies surrounding our own had a red shift in their hydrogen spectra that was proportional to their distance from us. Applying the Doppler Effect, Hubble recognized that this meant that all galaxies were receding from our own, with those farther away receding even faster. Knowing that our place in the universe was no more unique than any other, the implication was clear: The space within the universe itself was expanding. Just like pen marks on an expanding balloon, everything in the universe was accelerating away from everything else.

Figure 23.21 shows how the recession of galaxies looks like the remnants of a gigantic explosion, the famous **Big Bang**. Extrapolating backward in time, the Big Bang would have occurred between 13 and 15 billion years ago, when all matter would have been at a single point. From this, questions instantly arise. What caused the explosion? What happened before the Big Bang? Was there a before, or did time start then? For our purposes, the biggest question relating to the Big Bang is this: How does the Big Bang relate to the unification of the fundamental forces?

Figure 23.21 Galaxies are flying apart from one another, with the more distant ones moving faster, as if a primordial explosion expelled the matter from which they formed. The most distant known galaxies move nearly at the speed of light relative to us.

To fully understand the conditions of the very early universe, recognize that as the universe contracts to the size of the Big Bang, changes will occur. The density and temperature of the universe will increase dramatically. As particles become closer together, they will become too close to exist as we know them. The high energies will create other, more unusual particles to exist in greater abundance. Knowing this, let's move forward from the start of the universe, beginning with the Big Bang, as illustrated in Figure 23.22.

Figure 23.22 The evolution of the universe from the Big Bang onward (from left to right) is intimately tied to the laws of physics, especially those of particle physics at the earliest stages. Theories of the unification of forces at high energies may be verified by their shaping of the universe and its evolution.

The **Planck Epoch** $\left(0 \rightarrow 10^{-43}\text{ s}\right)$—Though scientists are unable to model the conditions of the Planck Epoch in the laboratory, speculation is that at this time compressed energy was great enough to reach the immense 10^{19} GeV necessary to unify gravity with all other forces. As a result, modern cosmology suggests that all four forces would have existed as one force, a hypothetical **superforce** as suggested by the **Theory of Everything**.

The **Grand Unification Epoch** $\left(10^{-43} \rightarrow 10^{-36}\text{ s}\right)$—As the universe expands, the temperatures necessary to maintain the superforce decrease. As a result, gravity separates, leaving the electroweak and strong nuclear forces together. At this time, the electromagnetic, weak, and strong forces are identical, matching the conditions requested in the Grand Unification Theory.

The **Inflationary Epoch** $\left(10^{-36} \rightarrow 10^{-32}\text{ s}\right)$—The separation of the strong nuclear force from the electroweak force during this time is thought to have been responsible for the massive inflation of the universe. Corresponding to the steep diagonal line on the left side of Figure 23.22, the universe may have expanded by a factor of 10^{50} or more in size. In fact, the expansion was so great during this time that it actually occurred faster than the speed of light! Unfortunately, there is little hope that we may be able to test the inflationary scenario directly since it occurs at energies near 10^{14} GeV, vastly greater than the limits of modern accelerators.

The **Electroweak Epoch** $\left(10^{-32} \rightarrow 10^{-11}\text{ s}\right)$—Now separated from both gravity and the strong nuclear force, the electroweak force exists as a singular force during this time period. As stated earlier, scientists are able to create the energies at this stage in the universe's expansion, needing only 100 GeV, as shown in Figure 23.20. W and Z bosons, as well as the Higgs boson, are released during this time.

The **Quark Era** $\left(10^{-11} \rightarrow 10^{-6}\text{ s}\right)$—During the Quark Era, the universe has expanded and temperatures have decreased to the

point at which all four fundamental forces have separated. Additionally, quarks began to take form as energies decreased.

As the universe expanded, further eras took place, allowing for the existence of hadrons, leptons, and photons, the fundamental particles of the standard model. Eventually, in nucleosynthesis, nuclei would be able to form, and the basic building blocks of atomic matter could take place. Using particle accelerators, we are very much working backwards in an attempt to understand the universe. It is encouraging to see that the macroscopic conditions of the Big Bang align nicely with our submicroscopic particle theory.

Check Your Understanding

19. Is there one grand unified theory or multiple grand unifying theories?
- a. one grand unifying theory
- b. multiple grand unifying theories

20. In what manner is $E = mc^2$ considered a precursor to the Grand Unified Theory?
- a. The grand unified theory seeks relate the electroweak and strong nuclear forces to one another just as $E = mc^2$ related energy and mass.
- b. The grand unified theory seeks to relate the electroweak force and mass to one another just as $E = mc^2$ related energy and mass.
- c. The grand unified theory seeks to relate the mass and strong nuclear forces to one another just as $E = mc^2$ related energy and mass.
- d. The grand unified theory seeks to relate gravity and strong nuclear force to one another, just as $E = mc^2$ related energy and mass.

21. List the following eras in order of occurrence from the Big Bang: Electroweak Epoch, Grand Unification Epoch, Inflationary Epoch, Planck Epoch, Quark Era.
- a. Quark Era, Grand Unification Epoch, Inflationary Epoch, Electroweak Epoch, Planck Epoch
- b. Planck Epoch, Inflationary Epoch, Grand Unification Epoch, Electroweak Epoch, Quark Era
- c. Planck Epoch, Electroweak Epoch, Grand Unification Epoch, Inflationary Epoch, Quark Era
- d. Planck Epoch, Grand Unification Epoch, Inflationary Epoch, Electroweak Epoch, Quark Era

22. How did the temperature of the universe change as it expanded?
- a. The temperature of the universe increased.
- b. The temperature of the universe decreased.
- c. The temperature of the universe first decreased and then increased.
- d. The temperature of the universe first increased and then decreased.

23. Under current conditions, is it possible for scientists to use particle accelerators to verify the Grand Unified Theory?
- a. No, there is not enough energy.
- b. Yes, there is enough energy.

24. Why are particles and antiparticles made to collide as shown in this image?

- a. Particles and antiparticles have the same mass.
- b. Particles and antiparticles have different mass.
- c. Particles and antiparticles have the same charge.
- d. Particles and antiparticles have opposite charges.

25. The existence of what particles were predicted as a consequence of the electroweak theory?
- a. fermions
- b. Higgs bosons

c. leptons
d. W^+, W^-, and Z^0 bosons

KEY TERMS

W^+ boson positive carrier particle of the weak nuclear force

W^- boson negative carrier particle of the weak nuclear force

Z^0 boson neutral carrier particle of the weak nuclear force

annihilation the process of destruction that occurs when a particle and antiparticle interact

antimatter matter constructed of antiparticles; antimatter shares most of the same properties of regular matter, with charge being the only difference between many particles and their antiparticle analogues

baryon hadrons that always decay to another baryon

Big Bang a gigantic explosion that threw out matter a few billion years ago

bottom quark a quark flavor

carrier particle a virtual particle exchanged in the transmission of a fundamental force

charmed quark a quark flavor, which is the counterpart of the strange quark

colliding beam head-on collisions between particles moving in opposite directions

color a property of quarks the relates to their interactions through the strong force

cyclotron accelerator that uses fixed-frequency alternating electric fields and fixed magnets to accelerate particles in a circular spiral path

down quark the second lightest of all quarks

Electroweak Epoch the stage before 10^{-11} back to 10^{-34} seconds after the Big Bang

electroweak theory theory showing connections between EM and weak forces

Feynman diagram a graph of time versus position that describes the exchange of virtual particles between subatomic particles

flavor quark type

gluons exchange particles of the nuclear strong force

Grand Unification Epoch the time period from 10^{-43} to 10^{-34} seconds after the Big Bang, when Grand Unification Theory, in which all forces except gravity are identical, governed the universe

Grand Unified Theory theory that shows unification of the strong and electroweak forces

graviton hypothesized particle exchanged between two particles of mass, transmitting the gravitational force between them

hadron particles composed of quarks that feel the strong and weak nuclear force

Higgs boson a massive particle that provides mass to the weak bosons and provides validity to the theory that carrier particles are identical under certain circumstances

Higgs field the field through which all fundamental particles travel that provides them varying mass through the transport of the Higgs boson

Inflationary Epoch the rapid expansion of the universe by an incredible factor of 10^{-50} for the brief time from 10^{-35} to about 10^{-32} seconds

lepton fundamental particles that do not feel the nuclear strong force

meson hadrons that can decay to leptons and leave no hadrons

pair production the creation of a particle and antiparticle, commonly an electron and positron, due to the annihilation of a photon

particle physics the study of and the quest for those truly fundamental particles having no substructure

pion particle exchanged between nucleons, transmitting the strong nuclear force between them

Planck Epoch the earliest era of the universe, before 10^{-43} seconds after the Big Bang

positron a particle of antimatter that has the properties of a positively charged electron

quantum chromodynamics the theory of color interaction between quarks that leads to understanding of the nuclear strong force

quantum electrodynamics the theory of electromagnetism on the particle scale

quark an elementary particle and fundamental constituent of matter that is a substructure of hadrons

Quark Era the time period from 10^{-11} to 10^{-6} seconds at which all four fundamental forces are separated and quarks begin to exit

Standard Model an organization of fundamental particles and forces that is a result of quantum chromodynamics and electroweak theory

strange quark the third lightest of all quarks

superforce the unification of all four fundamental forces into one force

synchrotron a version of a cyclotron in which the frequency of the alternating voltage and the magnetic field strength are increased as the beam particles are accelerated

Theory of Everything the theory that shows unification of all four fundamental forces

top quark a quark flavor

up quark the lightest of all quarks

weak nuclear force fundamental force responsible for particle decay

SECTION SUMMARY

23.1 The Four Fundamental Forces

- The four fundamental forces are gravity, the electromagnetic force, the weak nuclear force, and the strong nuclear force.
- A variety of particle accelerators have been used to explore the nature of subatomic particles and to test predictions of particle theories.

23.2 Quarks

- There are three types of fundamental particles—leptons, quarks, and carrier particles.
- Quarks come in six flavors and three colors and occur only in combinations that produce white.
- Hadrons are thought to be composed of quarks, with baryons having three quarks and mesons having a quark and an antiquark.
- Known particles can be divided into three major groups—leptons, hadrons, and carrier particles (gauge bosons).
- All particles of matter have an antimatter counterpart that has the opposite charge and certain other quantum numbers. These matter–antimatter pairs are otherwise very similar but will annihilate when brought together.
- The strong force is carried by eight proposed particles called gluons, which are intimately connected to a quantum number called color—their governing theory is thus called quantum chromodynamics (QCD). Taken together, QCD and the electroweak theory are widely accepted as the Standard Model of particle physics.

23.3 The Unification of Forces

- Attempts to show unification of the four forces are called Grand Unified Theories (GUTs) and have been partially successful, with connections proven between EM and weak forces in electroweak theory.
- Unification of the strong force is expected at such high energies that it cannot be directly tested, but it may have observable consequences in the as-yet-unobserved decay of the proton. Although unification of forces is generally anticipated, much remains to be done to prove its validity.

CHAPTER REVIEW

Concept Items

23.1 The Four Fundamental Forces

1. What forces does the inverse square law describe?
   a. the electromagnetic and weak nuclear force
   b. the electromagnetic force and strong nuclear force
   c. the electromagnetic force and gravity
   d. the strong nuclear force and gravity

2. Do the carrier particles explain the loss of mass in nuclear decay?
   a. no
   b. yes

3. What happens to the rate of voltage oscillation within a synchrotron each time the particle completes a loop?
   a. The rate of voltage oscillation increases as the particle travels faster and faster on each loop.
   b. The rate of voltage oscillation decreases as the particle travels faster and faster on each loop.
   c. The rate of voltage oscillation remains the same each time the particle completes a loop.
   d. The rate of voltage oscillation first increases and then remains constant each time the particle completes a loop.

4. Which of the four forces is responsible for ionic bonding?
   a. electromagnetic force
   b. gravity
   c. strong force
   d. weak nuclear force

5. What type of particle accelerator uses oscillating electric fields to accelerate particles around a fixed radius track?
   a. LINAC
   b. synchrotron
   c. SLAC
   d. Van de Graaff accelerator

23.2 Quarks

6. How does the charge of an individual quark determine hadron structure?
   a. Since the hadron must have an integral value, the individual quarks must be combined such that the average of their charges results in the value of a quark.
   b. Since the hadron must have an integral value, the individual atoms must be combined such that the sum of their charges is less than zero.
   c. The individual quarks must be combined such that the product of their charges is equal to the total charge of the hadron structure.
   d. Since the hadron must have an integral value of charge, the individual quarks must be combined such that the sum of their charges results in an

integral value.

7. Why do leptons not feel the strong nuclear force?
   a. Gluons are the carriers of the strong nuclear force that interacts between quarks through color interactions, but leptons are constructed of quarks that do not have gluons.
   b. Gluons are the carriers of the strong nuclear force that interacts between quarks through mass interactions, but leptons are not constructed of quarks and are not massive.
   c. Gluons are the carriers of the strong nuclear force that interacts between quarks through mass interactions, but leptons are constructed of the quarks that are not massive.
   d. Gluons are the carriers of the strong nuclear force that interacts between quarks through color interactions, but leptons are not constructed of quarks, nor do they have color constituents.

8. What property commonly distinguishes antimatter from its matter analogue?
   a. mass
   b. charge
   c. energy
   d. speed

9. Can the Standard Model change as new information is gathered?
   a. yes
   b. no

10. What is the relationship between the Higgs field and the Higgs boson?
   a. The Higgs boson is the carrier that transfers force for the Higgs field.
   b. The Higgs field is the time duration over which the Higgs particles transfer force to the other particles.
   c. The Higgs field is the magnitude of momentum transferred by the Higgs particles to the other particles.
   d. The Higgs field is the magnitude of torque transfers by the Higgs particles on the other particles.

11. What were the original three flavors of quarks discovered?
   a. up, down, and charm
   b. up, down, and bottom
   c. up, down, and strange
   d. up, down, and top

12. Protons are more massive than electrons. The three quarks in the proton account for only a small amount of this mass difference. What accounts for the remaining excess mass in protons compared to electrons?
   a. The highly energetic gluons connecting the quarks account for the remaining excess mass in protons compared to electrons.
   b. The highly energetic photons connecting the quarks account for the remaining excess mass in protons compared to electrons.
   c. The antiparallel orientation of the quarks present in a proton accounts for the remaining excess mass in protons compared to electrons.
   d. The parallel orientation of the quarks present in a proton accounts for the remaining excess mass in protons compared to electrons.

23.3 The Unification of Forces

13. Why is the unification of fundamental forces important?
   a. The unification of forces will help us understand fundamental structures of the universe.
   b. The unification of forces will help in the proof of the graviton.
   c. The unification of forces will help in achieving a speed greater than the speed of light.
   d. The unification of forces will help in studying antimatter particles.

14. Why are scientists unable to model the conditions of the universe at time periods shortly after the Big Bang?
   a. The amount of energy necessary to replicate the Planck Epoch is too high.
   b. The amount of energy necessary to replicate the Planck Epoch is too low.
   c. The volume of setup necessary to replicate the Planck Epoch is too high.
   d. The volume of setup necessary to replicate the Planck Epoch is too low.

15. What role does proton decay have in the search for GUTs?
   a. Proton decay is a premise of a number of GUTs.
   b. Proton decay negates the validity of a number of GUTs.

16. What is the name for the theory of unification of all four fundamental forces?
   a. the theory of everything
   b. the theory of energy-to-mass conversion
   c. the theory of relativity
   d. the theory of the Big Bang

17. Is it easier for scientists to find evidence for the Grand Unified Theory or the Theory of Everything? Explain.
   a. Theory of Everything, because it requires 10^{19} GeV of energy
   b. Theory of Everything, because it requires 10^{14} GeV of energy
   c. Grand Unified Theory, because it requires

10^{19} GeV of energy

d. Grand Unified Theory, because it requires 10^{14} GeV of energy

Critical Thinking Items

23.1 The Four Fundamental Forces

18. The gravitational force is considered a very weak force. Yet, it is strong enough to hold Earth in orbit around the Sun. Explain this apparent disparity.

a. At the level of the Earth-to-Sun distance, gravity is the strongest acting force because neither the strong nor the weak nuclear force exists at this distance.

b. At the level of the Earth-to-Sun distance, gravity is the strongest acting force because both the strong and the weak nuclear force is minimal at this distance.

19. True or False—Given that their carrier particles are massless, some may argue that the electromagnetic and gravitational forces should maintain the same value at all distances from their source. However, both forces decrease with distance at a rate of $\frac{1}{r}$

a. false

b. true

20. Why is a stationary target considered inefficient in a particle accelerator?

a. The stationary target recoils upon particle strike, thereby transferring much of the particle's energy into its motion. As a result, a greater amount of energy goes into breaking the particle into its constituent components.

b. The stationary target contains zero kinetic energy, so it requires more energy to break the particle into its constituent components.

c. The stationary target contains zero potential energy, so it requires more energy to break the particle into its constituent components.

d. The stationary target recoils upon particle strike, transferring much of the particle's energy into its motion. As a result, a lesser amount of energy goes into breaking the particle into its constituent components.

21. Compare the total strong nuclear force in a lithium atom to the total strong nuclear force in a lithium ion (Li^{+1}).

a. The total strong nuclear force in a lithium atom is thrice the total strong nuclear force in a lithium ion.

b. The total strong nuclear force in a lithium atom is twice the total strong nuclear force in a lithium ion.

c. The total strong nuclear force in a lithium atom is the same as the total strong nuclear force in a lithium ion.

d. The total strong nuclear force in a lithium atom is half the total strong nuclear force in a lithium ion.

23.2 Quarks

22. Explain why it is not possible to find a particle composed of just two quarks.

a. A particle composed of two quarks will have an integral charge and a white color. Hence, it cannot exist.

b. A particle composed of two quarks will have an integral charge and a color that is not white. Hence, it cannot exist.

c. A particle composed of two quarks will have a fractional charge and a white color. Hence, it cannot exist.

d. A particle composed of two quarks will have a fractional charge and a color that is not white. Hence, it cannot exist.

23. Why are mesons considered unstable?

a. Mesons are composites of two antiparticles that quickly annihilate each other.

b. Mesons are composites of two particles that quickly annihilate each other.

c. Mesons are composites of a particle and antiparticle that quickly annihilate each other.

d. Mesons are composites of two particles and one antiparticle that quickly annihilate each other.

24. Does antimatter have a negative mass?

a. No, antimatter does not have a negative mass.

b. Yes, antimatter does have a negative mass.

25. What similarities exist between the Standard Model and the periodic table of elements?

a. During their invention, both the Standard Model and the periodic table organized material by mass.

b. At the times of their invention, both the Standard Model and the periodic table organized material by charge.

c. At the times of their invention, both the Standard Model and the periodic table organized material by interaction with other available particles.

d. At the times of their invention, both the Standard Model and the periodic table organized material by size.

26. How were particle collisions used to provide evidence of the Higgs boson?

a. Because some particles do not contain the Higgs boson, the collisions of such particles will cause their destruction.
b. Because only the charged particles contain the Higgs boson, the collisions of such particles will cause their destruction and will expel the Higgs boson.
c. Because all particles with mass contain the Higgs boson, the collisions of such particles will cause their destruction and will absorb the Higgs boson.
d. Because all particles with mass contain the Higgs boson, the collisions of such particles will cause their destruction and will expel the Higgs boson.

27. Explain how the combination of a quark and antiquark can result in the creation of a hadron.
a. The combination of a quark and antiquark can result in a particle with an integer charge and color of white, therefore satisfying the properties for a hadron.
b. The combination of a quark and antiquark must result in a particle with a negative charge and color of white, therefore satisfying the properties for a hadron.
c. The combination of a quark and antiquark can result in a particle with an integer charge and color that is not white, therefore satisfying the properties for a hadron.
d. The combination of a quark and antiquark can result in particle with a fractional charge and color that is not white, therefore satisfying the properties for a hadron.

23.3 The Unification of Forces

28. Why does the strength of the strong force diminish under high-energy conditions?
a. Under high-energy conditions, particles interacting under the strong force will be compressed closer together. As a result, the force between them will decrease.
b. Under high-energy conditions, particles interacting under the strong force will start oscillating. As a result, the force between them will increase.
c. Under high-energy conditions, particles interacting under the strong force will have high velocity. As a result, the force between them will decrease.
d. Under high-energy conditions, particles interacting under the strong force will start moving randomly. As a result, the force between them will decrease.

29. If some unknown cause of the red shift, such as light becoming *tired* from traveling long distances through empty space, is discovered, what effect would there be on cosmology?
a. The effect would be substantial, as the Big Bang is based on the idea that the red shift is evidence that galaxies are moving toward one another.
b. The effect would be substantial, as the Big Bang is based on the idea that the red shift is evidence that the galaxies are moving away from one another.
c. The effect would be substantial, as the Big Bang is based on the idea that the red shift is evidence that galaxies are neither moving away from nor moving toward one another.
d. The effect would be substantial, as the Big Bang is based on the idea that the red shift is evidence that galaxies are sometimes moving away from and sometimes moving toward one another.

30. How many molecules of water are necessary if scientists wanted to check the 10^{31} -yr estimate of proton decay within the course of one calendar year?
a. 10^{29} molecules
b. 10^{30} molecules
c. 10^{31} molecules
d. 10^{32} molecules

31. As energy of interacting particles increases toward the theory of everything, the gravitational force between them increases. Why does this occur?
a. As energy increases, the masses of the interacting particles will increase.
b. As energy increases, the masses of the interacting particles will decrease.
c. As energy increases, the masses of the interacting particles will remain constant.
d. As energy increases, the masses of the interacting particles starts changing (increasing or decreasing). As a result, the gravitational force between the particles will increase.

Performance Task

23.3 The Unification of Forces

32. Communication is an often overlooked and useful skill for a scientist, especially in a competitive field where financial resources are limited. Scientists are often required to explain their findings or the relevance of their work to agencies within the government in order to maintain funding to continue their research. Let's say you are an ambitious young particle physicist, heading an expensive project, and you need to justify its existence to the appropriate funding agency. Write a brief paper (about one page) explaining why molecular-level structure is important in the functioning of designed materials in a specific industry.

- First, think of an industry where molecular-level structure is important.
- Research what materials are used in that industry as well as what are the desired properties of the materials.
- What molecular-level characteristics lead to what properties?

One example would be explaining how flexible but durable materials are made up of long-chained molecules and how this is useful for finding more environmentally friendly alternatives to plastics. Another example is explaining why electrically conductive materials are often made of metal and how this is useful for developing better batteries.

TEST PREP

Multiple Choice

23.1 The Four Fundamental Forces

33. Which of the following is not one of the four fundamental forces?
a. gravity
b. friction
c. strong nuclear
d. electromagnetic

34. What type of carrier particle has not yet been found?
a. gravitons
b. W bosons
c. Z bosons
d. pions

35. What effect does an increase in electric potential have on the accelerating capacity of a Van de Graaff generator?
a. It increases accelerating capacity.
b. It decreases accelerating capacity.
c. The accelerating capacity of a Van de Graaff generator is constant regardless of electric potential.
d. Van de Graaff generators do not have the capacity to accelerate particles.

36. What force or forces exist between a proton and a second proton?
a. The weak electrostatic force and strong magnetic force
b. The weak electrostatic and strong gravitational force
c. The weak frictional force and strong gravitational force
d. The weak nuclear force, the strong nuclear force, and the electromagnetic force

23.2 Quarks

37. To what color must quarks combine for a particle to be constructed?
a. black
b. green
c. red
d. white

38. What type of hadron is always constructed partially of an antiquark?
a. baryon
b. lepton
c. meson
d. photno

39. What particle is typically released when two particles annihilate?
a. graviton
b. antimatter
c. pion
d. photno

40. Which of the following categories is not one of the three main categories of the Standard Model?
a. gauge bosons
b. hadrons
c. leptons
d. quarks

41. Analysis of what particles began the search for the Higgs boson?
a. W and Z bosons
b. up and down quarks
c. mesons and baryons

d. neutrinos and photons

42. What similarities exist between the discovery of the quark and the discovery of the neutron?
a. Both the quark and the neutron were discovered by launching charged particles through an unknown structure and observing the particle recoil.
b. Both the quark and the neutron were discovered by launching electrically neutral particles through an unknown structure and observing the particle recoil.
c. Both quarks and neutrons were discovered by studying their deflection under an electric field.

23.3 The Unification of Forces

43. Which two forces were first combined, signifying the eventual desire for a Grand Unified Theory?
a. electric force and magnetic forces
b. electric force and weak nuclear force
c. gravitational force and the weak nuclear force
d. electroweak force and strong nuclear force

44. After the Big Bang, what was the first force to separate from the others?
a. electromagnetic force
b. gravity
c. strong nuclear force
d. weak nuclear force

45. What is the name of the device used by scientists to check for proton decay?
a. the cyclotron
b. the Large Hadron Collider
c. the Super-Kamiokande
d. the synchrotron

46. How do Feynman diagrams suggest the Grand Unified Theory?
a. The electromagnetic, weak, and strong nuclear forces all have similar Feynman diagrams.
b. The electromagnetic, weak, and gravitational forces all have similar Feynman diagrams.
c. The electromagnetic, weak, and strong forces all have different Feynman diagrams.

Short Answer

23.1 The Four Fundamental Forces

47. Why do people tend to be more aware of the gravitational and electromagnetic forces than the strong and weak nuclear forces?
a. The gravitational and electromagnetic forces act at short ranges, while strong and weak nuclear forces act at comparatively long range.
b. The strong and weak nuclear forces act at short ranges, while gravitational and electromagnetic forces act at comparatively long range.
c. The strong and weak nuclear forces act between all objects, while gravitational and electromagnetic forces act between smaller objects.
d. The strong and weak nuclear forces exist in outer space, while gravitational and electromagnetic forces exist everywhere.

48. What fundamental force is responsible for the force of friction?
a. the electromagnetic force
b. the strong nuclear force
c. the weak nuclear force

49. How do carrier particles relate to the concept of a force field?
a. Carrier particles carry mass from one location to another within a force field.
b. Carrier particles carry force from one location to another within a force field.
c. Carrier particles carry charge from one location to another within a force field.
d. Carrier particles carry volume from one location to another within a force field.

50. Which carrier particle is transmitted solely between nucleons?
a. graviton
b. photon
c. pion
d. W and Z bosons

51. Two particles of the same mass are traveling at the same speed but in opposite directions when they collide head-on.
What is the final kinetic energy of this two-particle system?
a. infinite
b. the sum of the kinetic energies of the two particles
c. zero
d. the product of the kinetic energies of the two particles

52. Why do colliding beams result in the location of smaller particles?
a. Colliding beams create energy, allowing more energy to be used to separate the colliding particles.
b. Colliding beams lower the energy of the system, so it requires less energy to separate the colliding particles.
c. Colliding beams reduce energy loss, so less energy

is required to separate colliding particles.
d. Colliding beams reduce energy loss, allowing more energy to be used to separate the colliding particles.

23.2 Quarks

53. What two features of quarks determine the structure of a particle?
a. the color and charge of individual quarks
b. the color and size of individual quarks
c. the charge and size of individual quarks
d. the charge and mass of individual quarks

54. What fundamental force does quantum chromodynamics describe?
a. the weak nuclear force
b. the strong nuclear force
c. the electromagnetic force
d. the gravitational force

55. Is it possible for a baryon to be constructed of two quarks and an antiquark?
a. Yes, the color of the three particles would be able to sum to white.
b. No, the color of the three particles would not be able to sum to white.

56. Can baryons be more massive than mesons?
a. no
b. yes

57. If antimatter exists, why is it so difficult to find?
a. There is a smaller amount of antimatter than matter in the universe; antimatter is quickly annihilated by its matter analogue.
b. There is a smaller amount of matter than antimatter in the universe; matter is annihilated by its antimatter analogue.
c. There is a smaller amount of antimatter than matter in universe; antimatter and its matter analogue coexist.
d. There is a smaller amount of matter than antimatter in the universe; matter and its antimatter analogue coexist.

58. Does a neutron have an antimatter counterpart?
a. No, the antineutron does not exist.
b. Yes, the antineutron does exist.

59. How are the four fundamental forces incorporated into the Standard Model of the atom?
a. The four fundamental forces are represented by their carrier particles, the electrons.
b. The four fundamental forces are represented by their carrier particles, the gauge bosons.
c. The four fundamental forces are represented by their carrier particles, the leptons.
d. The four fundamental forces are represented by their carrier particles, the quarks.

60. Which particles in the Standard Model account for the majority of matter with which we are familiar?
a. particles in fourth column of the Standard Model
b. particles in third column of the Standard Model
c. particles in the second column of the Standard Model
d. particles in the first column of the Standard Model

61. How can a particle gain mass by traveling through the Higgs field?
a. The Higgs field slows down passing particles; the decrease in kinetic energy is transferred to the particle's mass.
b. The Higgs field accelerates passing particles; the decrease in kinetic energy is transferred to the particle's mass.
c. The Higgs field slows down passing particles; the increase in kinetic energy is transferred to the particle's mass.
d. The Higgs field accelerates passing particles; the increase in kinetic energy is transferred to the particle's mass.

62. How does mass-energy conservation relate to the Higgs field?
a. The increase in a particle's energy when traveling through the Higgs field is countered by its increase in mass.
b. The decrease in a particle's kinetic energy when traveling through the Higgs field is countered by its increase in mass.
c. The decrease in a particle's energy when traveling through the Higgs field is countered by its decrease in mass.
d. The increase in a particle's energy when traveling through the Higgs field is countered by its decrease in mass.

23.3 The Unification of Forces

63. Why do scientists believe that the strong nuclear force and the electroweak force will combine under high energies?
a. The electroweak force will have greater strength.
b. The strong nuclear force and electroweak force will achieve the same strength.
c. The strong nuclear force will have greater strength.

64. At what energy will the strong nuclear force theoretically unite with the electroweak force?

a. 10^{12} eV
b. 10^{13} eV
c. 10^{14} eV
d. 10^{15} eV

65. While we can demonstrate the unification of certain forces within the laboratory, for how long were the four forces naturally unified within the universe?
a. 10^{-43} seconds
b. 10^{-41} seconds
c. 10^{-39} seconds
d. 10^{-38} seconds

66. How does the search for the Grand Unified Theory help test the standard cosmological model?
a. Scientists are increasing energy in the lab that models the energy in earlier, denser stages of the universe.
b. Scientists are increasing energy in the lab that models the energy in earlier, less dense stages of the universe.
c. Scientists are decreasing energy in the lab that models the energy in earlier, denser stages of the universe.
d. Scientists are decreasing energy in the lab that models the energy in earlier, less dense stages of the universe.

67. Why does finding proof that protons do not decay not disprove all GUTs?
a. Proton decay is not a premise of all GUTs, and current GUTs can be amended in response to new findings.
b. Proton decay is a premise of all GUTs, but current GUTs can be amended in response to new findings.

68. When accelerating elementary particles in a particle accelerator, they quickly achieve a speed approaching the speed of light. However, as time continues, the particles maintain this speed yet continue to increase their kinetic energy. How is this possible?
a. The speed remains the same, but the masses of the particles increase.
b. The speed remains the same, but the masses of the particles decrease.
c. The speed remains the same, and the masses of the particles remain the same.
d. The speed and masses will remain the same, but temperature will increase.

Extended Response

23.1 The Four Fundamental Forces

69. If the strong attractive force is the greatest of the four fundamental forces, are all masses fated to combine together at some point in the future? Explain.
a. No, the strong attractive force acts only at incredibly small distances. As a result, only masses close enough to be within its range will combine.
b. No, the strong attractive force acts only at large distances. As a result, only masses far enough apart will combine.
c. Yes, the strong attractive force acts at any distance. As a result, all masses are fated to combine together at some point in the future.
d. Yes, the strong attractive force acts at large distances. As a result, all masses are fated tocombine together at some point in the future.

70. How does the discussion of carrier particles relate to the concept of relativity?
a. Calculations of mass and energy during their transfer are relativistic, because carrier particles travel more slowly than the speed of sound.
b. Calculations of mass and energy during their transfer are relativistic, because carrier particles travel at or near the speed of light.
c. Calculations of mass and energy during their transfer are relativistic, because carrier particles travel at or near the speed of sound.
d. Calculations of mass and energy during their transfer are relativistic, because carrier particles travel faster than the speed of light.

71. Why are synchrotrons constructed to be very large?
a. By using a large radius, high particle velocities can be achieved using a large centripetal force created by large electromagnets.
b. By using a large radius, high particle velocities can be achieved without a large centripetal force created by large electromagnets.
c. By using a large radius, the velocities of particles can be reduced without a large centripetal force created by large electromagnets.
d. By using a large radius, the acceleration of particles can be decreased without a large centripetal force created by large electromagnets.

23.2 Quarks

72. In this image, how does the emission of the gluon cause the down quark to change from a red color to a green color?

a. The emitted red gluon is made up of a green and a red color. As a result, the down quark changes from a red color to a green color.
b. The emitted red gluon is made up of an anti-green and an anti-red color. As a result, the down quark changes from a red color to a green color.
c. The emitted red gluon is made up of a green and an anti-red color. As a result, the down quark changes from a red color to a green color.
d. The emitted red gluon is made up of an anti-green and a red color. As a result, the down quark changes from a red color to a green color.

73. Neutrinos are much more difficult for scientists to find when compared to other hadrons and leptons. Why is this?
a. Neutrinos are hadrons, and they lack charge.
b. Neutrinos are not hadrons, and they lack charge.
c. Neutrinos are hadrons, and they have positive charge.
d. Neutrinos are not hadrons, and they have a positive charge.

74. What happens to the masses of a particle and its antiparticle when the two annihilate at low energies?
a. The masses of the particle and antiparticle are transformed into energy in the form of photons.
b. The masses of the particle and antiparticle are converted into kinetic energy of the particle and antiparticle respectively.
c. The mass of the antiparticle is converted into kinetic energy of the particle.
d. The mass of the particle is converted into radiation energy of the antiparticle.

75. When a star erupts in a supernova explosion, huge numbers of electron neutrinos are formed in nuclear reactions. Such neutrinos from the 1987A supernova in the relatively nearby Magellanic Cloud were observed within hours of the initial brightening, indicating that they traveled to earth at approximately the speed of light. Explain how this data can be used to set an upper limit on the mass of the neutrino.
a. If the velocity of the neutrino is known, then the upper limit on mass of the neutrino can be set.
b. If only the kinetic energy of the neutrino is known, then the upper limit on mass of the neutrino can be set.
c. If either the velocity or the kinetic energy is known, then the upper limit on the mass of the neutrino can be set.
d. If both the kinetic energy and the velocity of the neutrino are known, then the upper limit on the mass of the neutrino can be set.

76. The term *force carrier particle* is shorthand for the scientific term *vector gauge boson*. From that perspective, can the Higgs boson truly be considered a *force carrier particle*?
a. No, the mass quality provided by the Higgs boson is a scalar quantity.
b. Yes, the mass quality provided by the Higgs boson results in a change of particle's direction.

23.3 The Unification of Forces

77. If a Grand Unified Theory is proven and the four forces are unified, it will still be correct to say that the orbit of the Moon is determined by the gravitational force. Explain why.
a. Gravity will not be a property of the unified force.
b. Gravity will be one property of the unified force.
c. Apart from gravity, no other force depends on the mass of the object.
d. Apart from gravity, no other force can make an object move in a fixed orbit.

78. As the universe expanded and temperatures dropped, the strong nuclear force separated from the electroweak force. Is it likely that under cooler conditions, the force of electricity will separate from the force of magnetism?
a. No, the electric force relies on the magnetic force and vice versa.
b. Yes, the electric and magnetic forces can be separated from each other.

79. Two pool balls collide head-on and stop. Their original kinetic energy is converted to heat and sound. Given that this is not possible for particles, what happens to their converted energy?
a. The kinetic energy is converted into relativistic potential energy, governed by the equation $E = \lambda mch$.
b. The kinetic energy is converted into relativistic mass, governed by the equation $E = \lambda m^2 c$.
c. The kinetic energy is converted into relativistic potential energy, governed by the equation $E = \lambda mgh$.

d. Their kinetic energy is converted into relativistic mass, governed by the equation $E = \lambda mc^2$.

APPENDIX A

Reference Tables

Figure A1 Periodic Table of Elements

Prefix	Symbol	Value	Prefix	Symbol	Value
tera	T	10^{12}	deci	d	10^{-1}
giga	G	10^{9}	centi	c	10^{-2}
mega	M	10^{6}	milli	m	10^{-3}
kilo	k	10^{3}	micro	μ	10^{-6}
hecto	h	10^{2}	nano	n	10^{-9}
deka	da	10^{1}	pico	p	10^{-12}

Table A1 Metric Prefixes for Powers of Ten and Their Symbols

Prefix	Symbol	Value	Prefix	Symbol	Value
—		10^0	femto	f	10^{-15}

Table A1 Metric Prefixes for Powers of Ten and Their Symbols

	Entity	Abbreviation	Name
Fundamental units	Length	m	meter
	Mass	kg	kilogram
	Time	s	second
	Current	A	ampere
Supplementary unit	Angle	rad	radian
Derived units	Force	$N = kg \cdot \frac{m}{s^2}$	newton
	Energy	$J = kg \cdot m^2$	joule
	Power	$W = \frac{J}{s}$	watt
	Pressure	$Pa = \frac{N}{m^2}$	pascal
	Frequency	$Hz = \frac{1}{s}$	hertz
	Electronic potential	$V = \frac{J}{C}$	volt
	Capacitance	$F = \frac{C}{V}$	farad
	Charge	$C = s \cdot A$	coulomb
	Resistance	$\Omega = \frac{V}{A}$	ohm
	Magnetic field	$T = \frac{N}{A \cdot m}$	tesla
	Nuclear decay rate	$Bq = \frac{1}{s}$	becquerel

Table A2 SI Units

Length	1 inch (in.) = 2.54 cm (exactly)
	1 foot (ft) = 0.3048 m
	1 mile (mi) = 1.609 km

Table A3 Selected British Units

Force	1 pound (lb) = 4.448 N
Energy	1 British thermal unit (Btu) = 1.055×10^3 J
Power	1 horsepower (hp) = 746 W
Pressure	1 lb/in^2 = 6.895×10^3 Pa

Table A3 Selected British Units

Length	1 light year (ly) = 9.46×10^{15} m
	1 astronomical unit (au) = 1.50×10^{11} m
	1 nautical mile = 1.852 km
	1 angstrom(Å) = 10^{-10} m
Area	1 acre (ac) = 4.05×10^3 m2
	1 square foot (ft^2) 9.29×10^{-2} m^3
	1 barn (*b*) = 10^{-28} m^2
Volume	1 liter (*L*) = 10^{-3} m^3
	1 U.S. gallon (gal) = 3.785×10^{-3} m^3
Mass	1 solar mass = 1.99×10^{30} kg
	1 metric ton = 10^3 kg
	1 atomic mass unit (*u*) = 1.6605×10^{-27} kg
Time	1 year (*y*) = 3.16×10^7 s
	1 day (*d*) = 86,400 s
Speed	1 mile per hour (mph) = 1.609 km / h
	1 nautical mile per hour (naut) = 1.852 km / h
Angle	1 degree (°) = 1.745×10^{-2} rad
	1 minute of arc (') = 1 / 60 degree
	1 second of arc (") = 1 / 60 minute of arc
	1 grad = 1.571×10^{-2} rad

Table A4 Other Units

Energy	1 kiloton TNT (kT) = 4.2×10^{12} J
	1 kilowatt hour (kW · *h*) = 3.60 × 106 J
	1 food calorie (kcal) = 4186 J
	1 calorie (cal) = 4.186 J
	1 electron volt (cV) = 1.60×10^{-19} J
Pressure	1 atmosphere (atm) = 1.013×10^{5} Pa
	1 millimeter of mercury (mm Hg) = 133.3 Pa
	1 torricelli (torr) = 1 mm Hg = 133.3 Pa
Nuclear decay rate	1 curie (Ci) = 3.70×10^{10} Bq

Table A4 Other Units

Circumference of a circle with radius *r* or diameter *d*	$C = 2\pi = \pi d$
Area of a circle with radius *r* or diameter *d*	$A = \pi r^2 = \pi d^2/4$
Area of a sphere with radius *r*	$A = 4\pi r^2$
Volume of a sphere with radius *r*	$V = (4/3)(\pi r^3)$

Table A5 Useful formulae

Symbol	Meaning	Best Value	Approximate Value
c	Speed of light in vacuum	2.99792458×10^{8} m/s	3.00×10^{8} m/s
G	Gravitational constant	$6.67384\,(80) \times 10^{-11}\,\mathrm{N \cdot m^2/kg^2}$	$6.67 \times 10^{-11}\,\mathrm{N \cdot m^2/kg^2}$
${}^{N}A$	Avogadro's number	$6.02214129\,(27) \times 10^{23}$ J/K	6.02×10^{23}
k	Boltzmann's constant	$1.3806488\,(13) \times 10^{-23}$ J/K	1.38×10^{-23} J/K
R	Gas constant	$8.3144621\,(75)$ J/mol · K	8.31 J/mol · K = 1.99 cal/mol · K = 0.0821 atm · L/mol · K
σ	Stefan-Boltzmann Constant	$5.670373\,(21) \times 10^{-8}\,\mathrm{W/m^2 \cdot K}$	$5.67 \times 10^{-8}\,\mathrm{W/m^2 \cdot K}$
k	Coulomb force constant	$8.987551788... \times 10^{9}\,\mathrm{N \cdot m^2/C^2}$	$8.99 \times 10^{9}\,\mathrm{N \cdot m^2/C^2}$

Table A6 Important Constants

Symbol	Meaning	Best Value	Approximate Value
q_e	Charge on electron	$-1.602176565\,(35) \times 10^{-19}\,\mathrm{C}$	$-1.60 \times 10^{-19}\,\mathrm{C}$
ε_0	Permittivity of free space	$8.854187817... \times 10^{-12}\,\mathrm{C^2/N \cdot m^2}$	$8.85 \times 10^{-12}\,\mathrm{C^2/N \cdot m^2}$
μ_0	Permeability of free space	$4\pi \times 10^{-7}\,\mathrm{T \cdot m/A}$	$1.26 \times 10^{-6}\,\mathrm{T \cdot m/A}$
h	Planck's constant	$6.62606957\,(29) \times 10^{-34}\,\mathrm{J \cdot s}$	$6.63 \times 10^{-34}\,\mathrm{J \cdot s}$

Table A6 Important Constants

Alpha	Α	α
Beta	Β	β
Gamma	Γ	γ
Delta	Δ	δ
Epsilon	Ε	ε
Zeta	Ζ	ζ
Eta	Η	η
Theta	Θ	θ
Iota	Ι	ι
Kappa	Κ	κ
Lambda	Λ	λ
Mu	Μ	μ
Nu	Ν	ν
Xi	Ξ	ξ
Omicron	Ο	ο
Pi	Π	π
Rho	Ρ	ρ
Sigma	Σ	σ
Tau	Τ	τ

Table A7 The Greek Alphabet

Upsilon	ϒ	υ
Phi	Φ	ϕ
Chi	X	χ
Psi	Ψ	ψ
Omega	Ω	ω

Table A7 The Greek Alphabet

Sun	mass	1.99×10^{30} kg
	average radius	6.96×10^{8} m
	Earth-sun distance (average)	1.496×10^{11} m
Earth	mass	5.9736×10^{24} kg
	average radius	6.376×10^{6} m
	orbital period	3.16×10^{7} s
Moon	mass	7.35×10^{22} kg
	average radius	1.74×10^{6} s
	orbital period (average)	2.36×10^{6} s
	Earth-moon distance (average)	3.84×10^{8} m

Table A8 Solar System Data

Atomic number, Z	Name	Atomic Mass Number, A	Symbol	Atomic Mass (u)	Percent Abundance or Decay Mode	Half-life, $^{t}1/2$
0	neutron	1	*n*	1.008 665	β^-	10.37 min
1	Hydrogen	1	^{1}H	1.007 825	99.985%	
	Deuterium	2	^{2}H or D	2.014 102	0.015%	
	Tritium	3	^{3}H or T	3.016 050	β^-	12.33 y
2	Helium	3	^{3}He	3.016 030	1.38×10^{-4} %	

Table A9 Atomic Masses and Decay

Atomic number, Z	Name	Atomic Mass Number, A	Symbol	Atomic Mass (u)	Percent Abundance or Decay Mode	Half-life, $^{t}1/2$
		4	^{4}He	4.002 603	≈100%	
3	Lithium	6	^{6}Li	6.015 121	7.5%	
		7	^{7}Li	7.016 003	92.5%	
4	Beryllium	7	^{7}Be	7.016 928	EC	53.29 d
		9	^{9}Be	9.012 182	100%	
5	Boron	10	^{10}B	10.012 937	19.9%	
		11	^{11}B	11.009 305	80.1%	
6	Carbon	11	^{11}C	11.011 432	EC, β^{+}	
		12	^{12}C	12.000 000	98.90%	
		13	^{13}C	13.003 355	1.10%	
		14	^{14}C	14.003 241	β^{-}	5730 y
7	Nitrogen	13	^{12}N	13.005 738	β^{+}	9.96 min
		14	^{13}N	14.003 074	99.63%	
		15	^{14}N	15.000 108	0.37%	
8	Oxygen	15	^{15}O	15.003 065	EC, β^{+}	122 s
		16	^{16}O	15.994 915	99.76%	
		18	^{18}O	17.999 160	0.200%	
9	Fluorine	18	^{18}F	18.000 937	EC, β^{+}	1.83 h
		19	^{19}F	18.998 403	100%	
10	Neon	20	^{20}Ne	19.992 435	90.51%	
		22	^{22}Ne	21.991 383	9.22%	
11	Sodium	22	^{22}Na	21.994 434	β^{+}	2.602 y
		23	^{23}Na	22.989 767	100%	
		24	^{24}Na	23.990 961	β^{-}	14.96 h

Table A9 Atomic Masses and Decay

Atomic number, Z	Name	Atomic Mass Number, A	Symbol	Atomic Mass (u)	Percent Abundance or Decay Mode	Half-life, $t_{1/2}$
12	Magnesium	24	^{24}Mg	23.985 042	78.99%	
13	Aluminum	27	^{27}Al	26.981 539	100%	
14	Silicon	28	^{28}Si	27.976 927	92.23%	2.62h
		31	^{31}Si	30.975 362	β^-	
15	Phosphorus	31	^{31}P	30.973 762	100%	
		32	^{32}P	31.973 907	β^-	14.28 d
16	Sulfur	32	^{32}S	31.972 070	95.02%	
		35	^{35}S	34.969 031	β^-	87.4 d
17	Chlorine	35	^{35}Cl	34.968 852	75.77%	
		37	^{37}Cl	36.965 903	24.23%	
18	Argon	40	^{40}Ar	39.962 384	99.60%	
19	Potassium	39	^{39}K	38.963 707	93.26%	
		40	^{40}K	39.963 999	0.0117%, EC, β^-	1.28×10^9 y
20	Calcium	40	^{40}Ca	39.962 591	96.94%	
21	Scandium	45	^{45}Sc	44.955 910	100%	
22	Titanium	48	^{48}Ti	47.947 947	73.8%	
23	Vanadium	51	^{51}V	50.943 962	99.75%	
24	Chromium	52	^{52}Cr	51.940 509	83.79%	
25	Manganese	55	^{55}Mn	54.938 047	100%	
26	Iron	56	^{56}Fe	55.934 939	91.72%	
27	Cobalt	59	^{59}Co	58.933 198	100%	
		60	^{60}Co	59.933 819	β^-	5.271 y
28	Nickel	58	^{58}Ni	57.935 346	68.27%	

Table A9 Atomic Masses and Decay

Atomic number, Z	Name	Atomic Mass Number, A	Symbol	Atomic Mass (u)	Percent Abundance or Decay Mode	Half-life, $t_{1/2}$
		60	^{60}Ni	59.930 788	26.10%	
29	Copper	63	^{63}Cu	62.939 598	69.17%	
			^{65}Cu	64.927 793	30.83%	
30	Zinc	64	^{64}Zn	63.929 145	48.6%	
		66	^{66}Zn	65.926 034	27.9%	
31	Gallium	69	^{69}Ga	68.925 580	60.1%	
32	Germanium	72	^{72}Ge	71.922 079	27.4%	
		74	^{74}Ge	73.921 177	36.5%	
33	Arsenic	75	^{75}As	74.921 594	100%	
34	Selenium	80	^{80}Se	79.916 520	49.7%	
35	Bromine	79	^{79}Br	78.918 336	50.69%	
36	Krypton	84	^{84}Kr	83.911 507	57.0%	
37	Rubidium	85	^{85}Rb	84.911 794	72.17%	
38	Strontium	86	^{86}Sr	85.909 267	9.86%	
		88	^{88}Sr	87.905 619	82.58%	
		90	^{90}Sr	89.907 738	β^-	28.8 y
39	Yttrium	89	^{89}Y	88.905 849	100%	
		90	^{90}Y	89.907 152	β^-	64.1 h
40	Zirconium	90	^{90}Zr	89.904 703	51.45%	
41	Niobium	93	^{93}Nb	92.906 377	100%	
42	Molybdenum	98	^{98}Mo	97.905 406	24.13%	
43	Technetium	98	^{98}Tc	97.907 215	β^-	4.2×10^6 y
44	Ruthenium	102	^{102}Ru	101.904 348	31.6%	

Table A9 Atomic Masses and Decay

Atomic number, Z	Name	Atomic Mass Number, A	Symbol	Atomic Mass (u)	Percent Abundance or Decay Mode	Half-life, $^{t}1/2$
45	Rhodium	103	^{103}Rh	102.905 500	100%	
46	Palladium	106	^{106}Pd	105.903 478	27.33%	
47	Silver	107	^{107}Ag	106.905 092	51.84%	
		109	^{109}Ag	108.904 757	48.16%	
48	Cadmium	114	^{114}Cd	113.903 357	28.73%	
49	Indium	115	^{115}In	114.903 880	95.7%, β^-	4.4×10^{14} y
50	Tin	120	^{120}Sn	119.902 200	32.59%	
51	Antimony	121	^{121}Sb	120.903 821	57.3%	
52	Tellurium	130	^{130}Te	129.906 229	33.8%, β^-	2.5×10^{21} y
53	Iodine	127	^{127}I	126.904 473	100%	
		131	^{131}I	130.906 114	β^-	8.040 d
54	Xenon	132	^{132}Xe	131.904 144	26.9%	
		136	^{136}Xe	135.907 214	8.9%	
55	Cesium	133	^{133}Cs	132.905 429	100%	
		134	^{134}Cs	133.906 696	EC, β^-	2.06 y
56	Barium	137	^{137}Ba	136.905 812	11.23%	
		138	^{138}Ba	137.905 232	71.70%	
57	Lanthanum	139	^{139}La	138.906 346	99.91%	
58	Cerium	140	^{140}Ce	139.905 433	88.48%	
59	Praseodymium	141	^{141}Pr	140.907 647	100%	
60	Neodymium	142	^{142}Nd	141.907 719	27.13%	
61	Promethium	145	^{145}Pm	144.912 743	EC, α	17.7 y
62	Samarium	152	^{152}Sm	151.919 729	26.7%	

Table A9 Atomic Masses and Decay

Atomic number, Z	Name	Atomic Mass Number, A	Symbol	Atomic Mass (u)	Percent Abundance or Decay Mode	Half-life, $^{t}1/2$
63	Europium	153	^{153}Eu	152.921 225	52.2%	
64	Gadolinium	158	^{158}Gd	157.924 099	24.84%	
65	Terbium	159	^{159}Tb	158.925 342	100%	
66	Dysprosium	164	^{164}Dy	163.929 171	28.2%	
67	Holmium	165	^{165}Ho	164.930 319	100%	
68	Erbium	166	^{166}Ho	165.930 290	33.6%	
69	Thulium	169	^{169}Tm	168.934 212	100%	
70	Ytterbium	174	^{174}Yb	173.938 859	31.8%	
71	Lutecium	175	^{175}Lu	174.940 770	97.41%	
72	Hafnium	180	^{180}Hf	179.946 545	35.10%	
73	Tantalum	181	^{181}Ta	180.947 992	99.98%	
74	Tungsten	184	^{184}W	183.950 928	30.67%	
75	Rhenium	187	^{187}Re	186.955 744	62.6%, β^-	4.6×10^{10}y
76	Osmium	191	^{191}Os	190.960 920	β^-	15.4 d
		192	^{192}Os	191.961 467	41.0%	
77	Iridium	191	^{191}Ir	190.960 584	37.3%	
		193	^{193}Ir	192.962 917	62.7%	
78	Platinum	195	^{195}Pt	194.964 766	33.8%	
79	Gold	197	^{197}Au	196.966 543	100%	
		198	^{198}Au	197.968 217	β^-	2.696 d
80	Mercury	199	^{199}Hg	198.968 253	16.87%	
		202	^{202}Hg	201.970 617	29.86%	
81	Thallium	205	^{205}Tl	204.974 401	70.48%	

Table A9 Atomic Masses and Decay

Atomic number, Z	Name	Atomic Mass Number, A	Symbol	Atomic Mass (u)	Percent Abundance or Decay Mode	Half-life, $t_{1/2}$
82	Lead	206	^{206}Pb	205.974 440	24.1%	
		207	^{207}Pb	206.975 872	22.1%	
		208	^{208}Pb	207.976 627	52.4%	
		210	^{210}Pb	209.984 163	α, β^-	22.3 y
		211	^{211}Pb	210.988 735	β^-	36.1 min
		212	^{212}Pb	211.991 871	β^-	10.64 h
83	Bismuth	209	^{209}Bi	208.980 374	100%	
		211	^{211}Bi	210.987 255	α, β^-	2.14 min
84	Polonium	210	^{210}Po	209.982 848	α	138.38 d
85	Astatine	218	^{218}At	218.008 684	α, β^-	1.6 s
86	Radon	222	^{222}Rn	222.017 570	α	3.82 d
87	Francium 2	223	^{223}Fr	223.019 733	α, β^-	21.8 min
88	Radium	226	^{226}Ra	226.025 402	α	1.60×10^3 y
89	Actinium	227	^{227}Ac	227.027 750	α, β^-	21.8 y
90	Thorium	228	^{228}Th	228.028 715	α	1.91 y
		232	^{232}Th	232.038 054	100%, α	1.41×10^{10} y
91	Protactinium	231	^{231}Pa	231.035 880	α	3.28×10^4 y
92	Uranium	233	^{233}U	233.039 628	α	1.59×10^3 y
		235	^{235}U	235.043 924	0.720%, α	7.04×10^8 y
		236	^{236}U	236.045 562	α	2.34×10^7 y

Table A9 Atomic Masses and Decay

Atomic number, Z	Name	Atomic Mass Number, A	Symbol	Atomic Mass (u)	Percent Abundance or Decay Mode	Half-life, $^{t}1/2$
		238	^{238}U	238.050 784	99.2745%, α	4.47×10^{9} y
		239	^{239}U	239.054 289	β^{-}	23.5 min
93	Neptunium	239	^{239}Np	239.052 933	β^{-}	2.355 d
94	Plutonium	239	^{239}Pu	239.052 157	α	2.41×10^{4} y
95	Americium	243	^{243}Am	243.061 375	α, fission	7.37×10^{3} y
96	Curium	245	^{245}Cm	245.065 483	α	8.50×10^{3} y
97	Berkelium	245	^{247}Bk	247.070 300	α	1.38×10^{3} y
98	Californium	249	^{249}Cf	249.074 844	α	351 y
99	Einsteinium	254	^{254}Es	254.088 019	α, β^{-}	276 d
100	Fermium	253	^{253}Fm	253.085 173	EC, α	3.00 d
101	Mendelevium	255	^{255}Md	255.091 081	EC, α	27 min
102	Nobelium	255	^{255}No	255.093 260	EC, α	3.1 min
103	Lawrencium	257	^{257}Lr	257.099 480	EC, α	0.646 s
104	Rutherfordium	261	^{261}Rf	261.108 690	α	1.08 mim
105	Dubnium	262	^{262}Db	262.113 760	α, fission	34 s
106	Seaborgium	263	^{263}Sg	263.11 86	α, fission	0.8 s
107	Bohrium	262	^{262}Bh	262.123 1	α	0.102 s
108	Hassium	264	^{264}Hs	264.128 5	α	0.08 ms
108	Meitnerium	266	^{266}Mt	266.137 8	α	3.4 ms

Table A9 Atomic Masses and Decay

Isotope	$^{t}1/2$	Decay Mode	Energy(MeV)	Percent		T-Ray Energy(MeV)	Percent
^{3}H	12.33 y	β^-	0.0186	100%			
^{14}C	5730 y	β^-	0.156	100%			
^{13}N	9.96 min	β^+	1.20	100%			
^{22}Na	2.602 y	β^+	1.20	90%	γ	1.27	100%
^{32}P	14.28 d	β^-	1.71	100%			
^{35}S	87.4 d	β^-	0.167	100%			
^{36}Ci	3.00 × 105 y	β^-	0.710	100%			
^{40}K	1.28 × 109 y	β^-	1.31	89%			
^{43}K	22.3 h	β^-	0.827	87%	γs	0.373	87%
						0.618	87%
^{45}Ca	165 d	β^-	0.257	100%			
^{51}Cr	27.70 d	EC			γ	0.320	10%
^{52}Mn	5.59d	β^+	3.69	28%	γ s	1.33	28%
						1.43	28%
^{52}Fe	8.27 h	β^+	1.80	43%		0.169	43%
						0.378	43%
^{59}Fe	44.6 d	β^- s	0.273	45%	γ s	1.10	57%
			0.466	55%		1.29	43%
^{60}Co	5.271 y	β^-	0.318	100%	γ s	1.17	100%
						1.33	100%
^{65}Zn	244.1 d	EC			γ	1.12	51%
^{67}Ga	78.3 h	EC			γ s	0.0933	70%
						0.185	35%
						0.300	19%

Table A10 Selected Radioactive Isotopes

Isotope	$^{t}1/2$	Decay Mode	Energy(MeV)	Percent		T-Ray Energy(MeV)	Percent
						others	
^{75}Se	118.5 d	EC			γ s	0.121	20%
						0.136	65%
						0.265	68%
						0.280	20%
						others	
^{86}Rb	18.8 d	β^- s	0.69	9%	γ	1.08	9%
			1.77	91%			
^{85}Sr	64.8 d	EC			γ	0.514 1	100%
^{90}Sr	28.8 y	β^-	0.546	100%			
^{90}Y	64.1 h	β^-	2.28	100%			
^{99m}Tc	6.02 h	IT			γ	0.142	100%
^{113}mIn	99.5 ^{m}in	IT			γ	0.392	100%
^{123}I	13.0 h	EC			γ	0.159	≈100%
^{131}I	8.040 d	β^- s	0.248	7%	γ s	0.364	85%
			0.607	93%		others	
			others				
^{129}Cs	32.3 h	EC			γ s	0.0400	35%
						0.372	32%
						0.411	25%
						others	
^{137}Cs	30.17 y	β^- s	0.511	95%	γ	0.662	95%
			1.17	5%			
^{140}Ba	12.79 d	β^-	1.035	≈100%	γ s	0.030	25%

Table A10 Selected Radioactive Isotopes

Isotope	$^{t}1/2$	Decay Mode	Energy(MeV)	Percent		T-Ray Energy(MeV)	Percent
						0.044	65%
						0.537	24%
						others	
^{198}Au	2.696 d	β^-	1.161	≈100%	γ	0.412	≈100%
^{197}Hg	64.1 h	EC			γ	0.0733	100%
^{210}Po	138.38 d	α	5.41	100%			
^{226}Ra	1.60 × 103 y	αs	4.68	5%	γ	0.186 1	100%
			4.87	95%			
^{235}U	7.038 × 108 y	α	4.68	≈100%	γ s	Numerous	<0.400%
^{238}U	4.468 × 109 y	αs	4.22	23%	γ	0.050	23%
			4.27	77%			
^{237}Np	2.14 × 106 y	αs	numerous		γ s	numerous	<0.250%
			4.96 (max.)				
^{239}Pu	2.41 × 104 y	αs	5.19	11%	γ s	7.5 × 10-5	73%
			5.23	15%		0.013	15%
			5.24	73%		0.052	15%
						others	
^{243}Am	7.37 × 103 y	αs	Max. 5.44		γ s	0.075	
			5.37	88%		others	
			5.32	11%			
			others				

Table A10 Selected Radioactive Isotopes

Symbol	Meaning	Best Value	Approximate Value
m_e	Electron mass	$9.10938291(40) \times 10^{-31}$ kg	9.11×10^{-31} kg

Table A11 Submicroscopic masses

Symbol	Meaning	Best Value	Approximate Value
m_p	Proton mass	$1.672621777(74) \times 10^{-27}$ kg	1.6726×10^{-27} kg
m_n	Neutron mass	$1.674927351(74) \times 10^{-27}$ kg	1.6749×10^{-27} kg
u	Atomic mass unit	$1.660538921(73) \times 10^{-27}$ kg	1.6605×10^{-27} kg

Table A11 Submicroscopic masses

Substance	ρ(kg/m^3)	Substance	ρ(kg/m^3)
Air	1.29	Iron	7.86×10^3
Air (at 20°C and Atmospheric pressure)	1.20	Lead	11.3×10^3
Aluminum	2.70×10^3	Mercury	13.6×10^3
Benzene	0.879×10^3	Nitrogen gas	1.25
Brass	8.4×10^3	Oak	0.710×10^3
Copper	8.92×10^3	Osmium	22.6×10^3
Ethyl alcohol	0.806×10^3	Oxygen gas	1.43
Fresh water	1.00×10^3	Pine	0.373×10^3
Glycerin	1.26×10^3	Platinum	21.4×10^3
Gold	1.93×10^3	Seawater	1.03×10^3
Helium gas	1.79×10^{-1}	Silver	10.5×10^3
Hydrogen gas	8.99×10^{-2}	Tin	7.30×10^3
Ice	0.917×10^3	Uranium	18.7×10^3

Table A12 Densities of common substances (including water at various temperatures)

Substance	Specific Heat (J/kg · °C)	Substance	Specific Heat (J/kg · °C)
Elemental solids		*Other solids*	
Aluminum	900	Brass	380
Beryllium	1830	Glass	837
Cadmium	230	Ice (−5 °C)	2090

Table A13 Specific heats of common substances

Substance	Specific Heat (J/kg · °C)	Substance	Specific Heat (J/kg · °C)
Copper	387	Marble	860
Germanium	322	Wood	1700
Gold	129	*Liquids*	
Iron	448	Alcohol (ethyl)	2400
Lead	128	Mercury	140
Silicon	703	Water (15 °C)	4186
Silver	234	*Gas*	
		Steam (100 °C)	2010

Note: To convert values to units of cal/g · °C, divide by 4186

Table A13 Specific heats of common substances

Substance	Melting Point (°C)	Latent Heat of Fusion (J/kg)	Boiling Point (°C)	Latent Heat of Vaporization (J/kg)
Helium	−272.2	5.23×10^3	−268.93	2.09×10^4
Oxygen	−218.79	1.38×10^4	−182.97	2.13×10^5
Nitrogen	−209.97	2.55×10^4	−195.81	2.01×10^5
Ethyl Alcohol	−114	1.04×10^5	78	8.54×10^5
Water	0.00	3.33×10^5	100.00	2.26×10^6
Sulfur	119	3.81×10^4	444.60	2.90×10^5
Lead	327.3	3.97×10^5	1750	8.70×10^5
Aluminum	660	3.97×10^5	2516	1.05×10^7
Silver	960.80	8.82×10^4	2162	2.33×10^6
Gold	1063.00	6.44×10^4	2856	1.58×10^6
Copper	1083	1.34×10^5	2562	5.06×10^6

Table A14 Heats of fusion and vaporization for common substances

Materials (Solids)	Average Linear Expansion Coefficient (α)(°C)$^{-1}$	Material (Liquids and Gases)	Average Volume Expansion Coefficient (B)(°C)$^{-1}$
Aluminum	24×10^{-6}	Acetone	1.5×10^{-4}
Brass and Bronze	19×10^{-6}	Alcohol, ethyl	1.12×10^{-4}
Concrete	12×10^{-6}	Benzene	1.24×10^{-4}
Copper	17×10^{-6}	Gasoline	9.6×10^{-4}
Glass (ordinary)	9×10^{-6}	Glycerin	4.85×10^{-4}
Glass (Pyrex)	3.2×10^{-6}	Mercury	1.82×10^{-4}
Invar (Ni-Fe alloy)	1.3×10^{-6}	Turpentine	9.0×10^{-4}
Lead	29×10^{-6}	Air* at 0°C	3.67×10^{-3}
Steel	13×10^{-6}	Helium*	3.665×10^{-3}

* The values given here assume the gases undergo expansion at constant pressure. However, the expansion of gases depends on the pressure applied to the gas. Therefore, gases do not have a specific value for the volume expansion coefficient.

Table A15 Coefficients of thermal expansion for common substances

Medium	v(m/s)	Medium	v(m/s)	Medium	v(m/s)
Gases		Liquids at 25°C		Solids*	
Hydrogen	1286	Glycerol	1904	Pyrex glass	5640
Helium	972	Seawater	1533	Iron	5950
Air	343	Water	1493	Aluminum	5100
Air	331	Mercury	1450	Brass	4700
Oxygen	317	Kerosene	1324	Copper	3560
		Methyl Alcohol	1143	Gold	3240
		Carbon tetrachloride	926	Lucite	2680
				Lead	1322
				Rubber	1600

Table A16 Speed of sound in various substances

Medium	*v*(m/s)	Medium	*v*(m/s)	Medium	*v*(m/s)

*Values given here are for propagation of longitudinal waves in bulk media. However, speeds for longitudinal waves in thin rods are slower, and speeds of transverse waves in bulk are even slower.

Table A16 Speed of sound in various substances

Source of Sound	*B(dB)*
Nearby jet airplane	150
Jackhammer machine gun	130
Siren; rock concert	120
Subway; power lawn mower	100
Busy traffic	80
Vacuum cleaner	70
Normal Conversation	60
Mosquito buzzing	40
whisper	30
Rustling leaves	10
Threshold of hearing	0

Table A17 Conversion of sound intensity to decibel level

Wavelength Range (nm)	Color Description
400-430	Violet
430-485	Blue
485-560	Green
560-590	Yellow
590-625	Orange
625-700	Red

Table A18 Wavelengths of visible light

Substance	Index of Refraction	Substance	Index of Refraction
Solids at 20°C		Liquids at 20°C	
Cubic zirconia	2.15	Benzene	1.501
Diamond (C)	2.419	Carbon disulfide	1.628
Flourite (CaF_2)	1.434	Carbon tetrachloride	1.461
Fused quartz (SiO_2)	1.458	Ethyl alcohol	1.361
Gallium phosphide	3.50	Glycerin	1.473
Glass, crown	1.52	Water	1.333
Glass, flint	1.66		
Ice (H_2O)	1.309	Gases at 0°C, 1 *atm*	
Polystyrene	1.49	Air	1.000 293
Sodium chloride (NaCl)	1.544	Carbon dioxide	1.000 45

Note: These values assume that light has a wavelength of 589 nm in vacuum.

Table A19 Indices of refraction

Hoop or thin cylindrical shell	$I_{CM} = MR^2$
Hollow cylinder	$I_{CM} = \frac{1}{2}M(R_1^2 + R_2^2)$
Solid cylinder or disk	$I_{CM} = \frac{1}{2}MR^2$
Rectangular plane	$I_{CM} = \frac{1}{12}M(a^2 + b^2)$
Long, thin rod with rotation axis through center	$I_{CM} = \frac{1}{12}ML^2$
Long, thin rod with rotation axis through end	$I_{CM} = \frac{1}{3}ML^2$
Solid sphere	$I_{CM} = \frac{2}{5}MR^2$
Thin spherical shell	$I_{CM} = \frac{2}{3}MR^2$

Table A20 Moments of inertia for different shapes

	μ_s	μ_k
Rubber on dry concrete	1.0	0.8

Table A21 Coefficients of friction for common objects on other objects

	μ_s	μ_k
Steel on steel	0.74	0.57
Aluminum on steel	0.61	0.47
Glass on glass	0.94	0.4
Copper on steel	0.53	0.36
Wood on wood	0.25-0.5	0.2
Waxed wood on wet snow	0.14	0.1
Waxed wood on dry snow	0.1	0.04
Metal on metal (lubricated)	0.15	0.06
Teflon on Teflon	0.04	0.04
Ice on ice	0.1	0.03
Synovial joints in humans	0.01	0.003

Note: All values are approximate. In some cases, the coefficient of friction can exceed 1.0.

Table A21 Coefficients of friction for common objects on other objects

Material	Dielectric Constant κ	Dielectric Strength* (10^6V/m)
Air (dry)	1.000 59	3
Bakelite	4.9	24
Fused quartz	4.3	8
Mylar	3.2	7
Neoprene rubber	6.7	12
Nylon	3.4	14
Paper	3.7	16
Paraffin-impregnated paper	3.5	11
Polystyrene	2.56	24
Polyvinyl chloride	3.4	40
Porcelain	6	8

Table A22 Dielectric constants

Material	Dielectric Constant κ	Dielectric Strength* (10^6V/m)
Pyrex glass	5.6	14
Silicone oil	2.5	15
Strontium titanate	233	8
Teflon	2.1	60
Vacuum	1.000 00	∞
Water	80	3

Table A22 Dielectric constants

INDEX

A

aberration 505
absolute zero 329
absorption spectrum 725
Acceleration 73, 93, 93, 94, 104, 116, 122, 128, 205
acceleration due to gravity 104
Accuracy 26
acoustics 8
activity 744
air resistance 162
alpha (α) decay 738
alpha particle 738
alternating current 606
ampere 604
amplitude 180, 390, 417, 457
amplitude modulation 461
analytical method 153
Anger camera 758
angle of incidence 478
angle of reflection 478
angle of refraction 491
angle of rotation 198
angular acceleration 212, 261
angular momentum 261
angular velocity 198, 206
annihilation 753, 785
Antimatter 784
antinode 403, 437
Antoine Henri Becquerel 736
aphelion 230
arc length 198
atomic number 735
atoms 8
Average acceleration 94
Average speed 62
Average velocity 63

B

baryons 784
beat frequency 436
Beats 436
becquerel 744
beta (β β) decay 738
Big Bang 792
binding energy 317, 699
blackbody 692
Boltzmann constant 359
bottom 781

C

capacitor 582
carbon-14 dating 745
carrier particles 773
Celsius scale 328
central axis 481
centrifugal force 205
centripetal acceleration 205, 213
centripetal force 207
cesium atomic clock 20
chain reaction 749
change in momentum 254
charmed 781
chromatic aberration 505
circuit diagrams 613
Circular motion 198, 205
classical physics 8
closed system 286
closed-pipe resonator 439
coefficient of friction 119
Colliding beams 778
collision 256
collisions 262
color 782
complex machine 293
component 154
Compton 705
Compton effect 705
concave lens 499
concave mirror 481
Condensation 341
conduction 334, 556
conductors 555
conservation 259
Constant acceleration 99
constructive interference 400
convection 334
conventional current 605
converging lens 498
conversion factor 24, 24
convex lens 498
convex mirror 481
Copernican model 232
corner reflector 494
Coulomb 562
coulomb force 772
Coulomb's constant 563
Coulomb's law 563
critical angle 493
critical mass 749
Curie temperature 654
cyclical process 373
cyclotron 777

D

damping 435
decay constant 744
decibels 424
deformation 178
degree Celsius (°C (°C) 328
degree Fahrenheit (°F (°F) 328
dependent variable 34, 67
derived units 19
destructive interference 401
dielectric 585
differential interference contrast (DIC) 535
diffraction 525
diffraction grating 533
diffused 479
direct current 606
direction 54

dispersion 491
displacement 57, 73, 94, 100
distance 56
diverging lens 499
domains 653
Doppler effect 430
down 780
dynamics 7, 116

E

E.O. Lawrence 777
eccentricity 234
Edwin Hubble 792
efficiency output 293
Einstein 698
elastic collision 262
electric circuits 612
Electric current 604
electric eye 702
electric field 456, 567
Electric motors 665
electric potential 575
Electric potential energy 572
electric power 632
electrical charge 562
Electricity 8
electromagnet 656
electromagnetic force 772
electromagnetic radiation (EMR) 456
electromagnetism 661
electron 551
electron cloud model 15
Electroweak Epoch 793
electroweak theory 790
emf 673
emission spectrum 724
energy 280, 390
energy-level diagram 726
energy-mass curve 751
English units 19
Enrico Fermi 773
entropy 366
equilibrium position 180
equivalent resistor 618
Ernest Rutherford 722
ether 306
excited states 726
experiment 14
exponential relationship 37
external force 116
eyepiece 504

F

Fahrenheit scale 328
Faraday 567
Fermi National Accelerator Laboratory 778
ferromagnetic 653
Feynman diagram 774
first law of thermodynamics 361
flavors 780
Fletcher 551
focal length 481
focal point 481
Force 116, 254
force field 567
frame of reference 306
frame-dragging effect 242
frames of reference 314
Fraunhofer lines 725
free-body diagram 116
freefall 124
Freezing 341
Frequency 180, 390, 416, 457
frequency modulation 461
friction 118
fulcrum 215, 290
fundamental 438
fundamental physical 18
fusion reactors 756

G

Galilei 115
Galileo 55
gamma (γ) rays 457
Gamma decay 739
gauge pressure 417
Geiger tube 742
General relativity 305
generator 666
geodetic effect 242
geometric optics 478
George Zweig 780
glass rods 550
gluon 782
gluons 774
Grand Unification Epoch 793
Grand Unified Theory (GUT) 790
graph 67, 73
graphical 144
gravitational 129
gravitational constant 238
gravitational force 117, 123, 772
Gravitational potential energy 280
graviton 774, 776
Gravity 104, 123
ground state 726
gun-type bomb 754

H

hadrons 780
half-life 743
harmonic motion 416
harmonics 438
head 144
head-to-tail method 144
Hearing 427
Heat 328
heat capacity 332
heat engine 372
Heat pumps 374
Heisenberg uncertainty principle 733
hertz 180
Hideki Yukawa 773
Higgs boson 785
Higgs field 787
Hooke's law 179
Huygens's principle 525

hydrogen-like atoms 728
hypothesis 14

I

ideal gas law 358
ideal mechanical advantage 290
illuminance 467
impulse 254
impulse-momentum theorem 254
in parallel 621
in series 618
incident ray 491
inclined plane 291
independent 162
independent variable 34, 67
index of refraction 488
induction 557, 672
inelastic collision 264
Inertia 120
inertial reference frame 307
Inflationary Epoch 793
infrared (IR) radiation 457
infrasound 427
input work 293
instantaneous acceleration 94
instantaneous speed 62
instantaneous velocity 64
insulator 555
interference 463
Internal energy 361
Inverse proportionality 37
inverse relationship 37
inverse-square law 564
Inversely proportional 123
inversion 403
ionizing radiation 759
iridescence 533
isolated system 260
isotope 736

J

joule 281

K

kelvin 20, 329
Kelvin scale 329
Kepler 115
Kepler's laws of planetary motion 230
kinematic equations 99
kinematics 7, 54, 67, 116
kinematics of rotational motion 214
kinetic 316
kinetic energy 262, 280, 572
Kinetic friction 171

L

laminated-coil transformer 669
Large Hadron Collider 776
laser 532
Laser Interferometer Gravitational-Wave Observatory 776
latent heat 341
latent heat of fusion 342
latent heat of vaporization 342
law of conservation of charge 554
law of conservation of energy 286
law of conservation of momentum 260, 262
law of inertia 120
law of reflection 478
law of refraction 478
length contraction 314
Leptons 780
lever 290
lever arm 216
light years 466
LightSail-1 708
line graph 34
linear acceleration 213
Linear momentum 254
linear relationships 37
liquid drop model 748
log-log plot 38
logarithmic (log) scale 37
longitudinal wave 391
longitudinal waves 416
loudness 427
Louis de Broglie 732
lumens 466
luminous flux 466
lux 467

M

magnetic dipoles 652
magnetic field 456, 654
magnetic flux 675
magnetic induction 672
magnetic pole 650
Magnetism 8
magnetized 653
magnitude 58
Manhattan Project 754
Mass 120, 122, 254
mass defect 317
mass number 735
maximum height 166
maximum height of a projectile 168
Maxwell's equations 458
mechanical advantage 290
mechanical energy 280
mechanical waves 390
medium 390
Melting 341
meson 774
mesons 784
method of adding percents 29
metric system 21
metric units 21
Michelson–Morley experiment 306
microwave radiation 457
Millikan 551

model 14
modern physics 8
moment of inertia 261
momentum 259, 262, 315
monochromatic 528
monochromators 536
motion 54, 62
Murray Gell-Mann 780

N

natural frequency 435
negative acceleration 94
negative charges 550
net external force 122, 128
net force 116, 128, 129
neutrino 739, 782
Newton 115, 123
Newton's first law of motion 118
Newton's laws of motion 115, 116
Newton's second law of motion 16, 122
Newton's third law of motion 128
Newton's universal law of gravitation 238
Niels Bohr 725
node 438
nodes 403
nonohmic 611
normal force 117, 119, 129
north pole 650
nuclear fission 748
Nuclear fusion 751
Nuclear physics 8
nuclear strong force 751
nucleons 735, 773
nuclide 736

O

objectives 504
observation 14
ocular 504
Ohm's law 609
ohmic 609
oil-drop 552
open-pipe resonator 439
optics 8
order of magnitude 24
oscillate 179
output work 293
overtones 438

P

pair production 785
parfocal 504
Particle physics 771
particle-wave duality 708
Pascal's principle 16
perihelion 230
period 180, 232, 390
Periodic motion 180
periodic wave 390
permanent magnet 653
permittivity of free space 583
Peter Higgs 787
phase change 341
phase diagram 343
photoelectric effect 698
photoelectron 698
photon 774
photon momentum 705
photons 698
photovoltaic cells 702
physical models 15
physical science 161
physics 6
pigment 460
pion 773
pitch 427
pivot point 215
Planck 693
Planck Epoch 793
Planck's constant 694
planetary model of the atom 723
plasma 341
plum pudding model 722
point charge 567
point masses 265
polarization 557
polarized light 464
position 54, 67, 73
positive charges 550
positron 784
Positron emission tomography 759
postulates 307
potential difference 575
potential energy 280, 316, 572
Power 281, 423
Precision 27
pressure 358
primary and secondary coils 669
principles 16
projectile 162
Projectile motion 162
Proper length 314
proportionality 122
proton decay 791
proton-proton cycle 753
protons 551
Ptolemaic model 232
pulley 292
pulse wave 390

Q

quadratic relationship 37
quantized 695
quantum 694
quantum chromodynamics 774, 782
quantum electrodynamics 774
quantum mechanics 8, 694
Quark Era 793
quarks 780

R

rad 760

Radar 461
radian 199
radiation 334
radio waves 457
radioactive 738
Radioactive dating 745
radioactive decay 738
Radioactivity 738
radiopharmaceutical 758
Radiotherapy 760
radius of curvature 198
rarefactions 416
rate 62
ray 478
Rayleigh criterion 538
real image 479
reciprocal 283
recoil 263
reference frame 54
reflection 403, 478
refracted ray 491
refraction 404
relative biological effectiveness 761
relative speed 314
Relativistic 314
relativistic energy 316
relativistic factor 313
Relativistic momentum 315
Relativity 305
Resistance 608
resistor 613
resolution 537
resonance 435
resonate 435
rest mass 315
restoring force 179
resultant 145
resultant vector 147
revolution 199
Richard Feynman 774
right-hand rule 656
roentgen equivalent man 761
rotational inertia 261
Rotational motion 198
Rutherford 551
Rutherford scattering 723
Rydberg constant 731

S

scalar 58, 62, 254
scientific law 16
scientific method 14
Scientific models 15
Scientific notation 23
scintillators 742
screw 292
second law of thermodynamics 366
semi-log plot 38
shock wave 432
SI units 19
significant figures 30
simple harmonic motion 179
Simple machines 290
simple pendulum 181
Simultaneity 310
Single-photon-emission computer tomography 758
slope 36
Snap Lab 54
Snell's law 491
solenoid 657
sonic boom 432
Sound 416
sound intensity 423
sound intensity level 424
south pole 650
special relativity 305, 310
specific heat 332
specular 479
speed 54, 62, 122, 417
Spin 198
Standard Model 780
standing waves 402
Stanford Linear Accelerator Center 778
static friction 171
statics 7
stationary 54
steady state 616
strange 780
strong nuclear force 737, 772
Sublimation 341
Super-Kamiokande 791
superforce 793
superposition 400
supersonic 432
synchrotron 777
system 119
systems 116

T

tagged 758
tail 144
tangent 71
tangential acceleration 213
Tangential velocity 200, 213
temperature 328
tension 129
test charge 567
The net external force 116
theory 16
Theory of Everything 793
theory of relativity 9
therapeutic ratio 760
thermal efficiency 376
thermal energy 328
thermal equilibrium 356
thermodynamics 8
thermonuclear bomb 755
thin-film interference 463
Thomson 550
three-dimensional models 15
thrust 129
time 62, 67, 73, 94, 95
Time dilation 313
top 781
torque 216, 261
torsion balance 562
total energy 316
total internal reflection 493
trajectory 162
transformer 669
translational acceleration 261
transmutation 738
transverse wave 391

transverse waves 416
trend line 36
twin paradox 314

U

U.S. customary units 21
ultrasound 427
ultraviolet (UV) radiation 457
ultraviolet catastrophe 693
uncertainty 28
uniform circular motion 205
uniform electric field 580
universal 16
universal laws 116
up 780

V

Van de Graaff 558
Van de Graaff generator 777
Vaporization 341
variables 34
vector 58, 63, 123, 254
vector addition 144
vector quantities 95
Vector subtraction 146
Velocity 63, 73, 93, 94, 99, 116, 122, 205, 254
virtual image 479
visible light 456
voltage 575

W

W bosons 775
watts 282
wave 390
wave velocity 394
wavefronts 525
wavelength 394, 419, 457
weak nuclear force 772
wedge 291
weight 119, 123, 129
Werner Heisenberg 733
wheel and axle 291
work 280
work–energy theorem 281

X

X-rays 457

Z

Z bosons 775
zeroth law of thermodynamics 356

FHSST Authors

The Free High School Science Texts: Textbooks for High School Students Studying the Sciences Chemistry Grades 10 - 12

Version 0
November 9, 2008

Copyright 2007 "Free High School Science Texts"
Permission **is** granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License".

STOP!!!!

Did you notice the **FREEDOMS** we've granted you?

Our copyright license is **different!** It grants freedoms rather than just imposing restrictions like all those other textbooks you probably own or use.

- We know people copy textbooks illegally but we would LOVE it if you copied our's - go ahead copy to your hearts content, **legally!**
- Publishers' revenue is generated by controlling the market, we don't want any money, go ahead, distribute our books far and wide - we DARE you!
- Ever wanted to change your textbook? Of course you have! Go ahead, change ours, make your own version, get your friends together, rip it apart and put it back together the way you like it. That's what we really want!
- Copy, modify, adapt, enhance, share, critique, adore, and contextualise. Do it all, do it with your colleagues, your friends, or alone but get involved! Together we can overcome the challenges our complex and diverse country presents.
- So what is the catch? The only thing you can't do is take this book, make a few changes and then tell others that they can't do the same with your changes. It's share and share-alike and we know you'll agree that is only fair.
- These books were written by volunteers who want to help support education, who want the facts to be freely available for teachers to copy, adapt and re-use. Thousands of hours went into making them and they are a gift to everyone in the education community.

FHSST Core Team

Mark Horner ; Samuel Halliday ; Sarah Blyth ; Rory Adams ; Spencer Wheaton

FHSST Editors

Jaynie Padayachee ; Joanne Boulle ; Diana Mulcahy ; Annette Nell ; René Toerien ; Donovan Whitfield

FHSST Contributors

Rory Adams ; Prashant Arora ; Richard Baxter ; Dr. Sarah Blyth ; Sebastian Bodenstein ; Graeme Broster ; Richard Case ; Brett Cocks ; Tim Crombie ; Dr. Anne Dabrowski ; Laura Daniels ; Sean Dobbs ; Fernando Durrell ; Dr. Dan Dwyer ; Frans van Eeden ; Giovanni Franzoni ; Ingrid von Glehn ; Tamara von Glehn ; Lindsay Glesener ; Dr. Vanessa Godfrey ; Dr. Johan Gonzalez ; Hemant Gopal ; Umeshree Govender ; Heather Gray ; Lynn Greeff ; Dr. Tom Gutierrez ; Brooke Haag ; Kate Hadley ; Dr. Sam Halliday ; Asheena Hanuman ; Neil Hart ; Nicholas Hatcher ; Dr. Mark Horner ; Robert Hovden ; Mfandaidza Hove ; Jennifer Hsieh ; Clare Johnson ; Luke Jordan ; Tana Joseph ; Dr. Jennifer Klay ; Lara Kruger ; Sihle Kubheka ; Andrew Kubik ; Dr. Marco van Leeuwen ; Dr. Anton Machacek ; Dr. Komal Maheshwari ; Kosma von Maltitz ; Nicole Masureik ; John Mathew ; JoEllen McBride ; Nikolai Meures ; Riana Meyer ; Jenny Miller ; Abdul Mirza ; Asogan Moodaly ; Jothi Moodley ; Nolene Naidu ; Tyrone Negus ; Thomas O'Donnell ; Dr. Markus Oldenburg ; Dr. Jaynie Padayachee ; Nicolette Pekeur ; Sirika Pillay ; Jacques Plaut ; Andrea Prinsloo ; Joseph Raimondo ; Sanya Rajani ; Prof. Sergey Rakityansky ; Alastair Ramlakan ; Razvan Remsing ; Max Richter ; Sean Riddle ; Evan Robinson ; Dr. Andrew Rose ; Bianca Ruddy ; Katie Russell ; Duncan Scott ; Helen Seals ; Ian Sherratt ; Roger Sieloff ; Bradley Smith ; Greg Solomon ; Mike Stringer ; Shen Tian ; Robert Torregrosa ; Jimmy Tseng ; Helen Waugh ; Dr. Dawn Webber ; Michelle Wen ; Dr. Alexander Wetzler ; Dr. Spencer Wheaton ; Vivian White ; Dr. Gerald Wigger ; Harry Wiggins ; Wendy Williams ; Julie Wilson ; Andrew Wood ; Emma Wormauld ; Sahal Yacoob ; Jean Youssef

Contributors and editors have made a sincere effort to produce an accurate and useful resource. Should you have suggestions, find mistakes or be prepared to donate material for inclusion, please don't hesitate to contact us. We intend to work with all who are willing to help make this a continuously evolving resource!

www.fhsst.org

Contents

I Introduction **1**

II Matter and Materials **3**

1 Classification of Matter - Grade 10 **5**

1.1 Mixtures . . . 5

1.1.1 Heterogeneous mixtures . . . 6

1.1.2 Homogeneous mixtures . . . 6

1.1.3 Separating mixtures . . . 7

1.2 Pure Substances: Elements and Compounds . . . 9

1.2.1 Elements . . . 9

1.2.2 Compounds . . . 9

1.3 Giving names and formulae to substances . . . 10

1.4 Metals, Semi-metals and Non-metals . . . 13

1.4.1 Metals . . . 13

1.4.2 Non-metals . . . 14

1.4.3 Semi-metals . . . 14

1.5 Electrical conductors, semi-conductors and insulators . . . 14

1.6 Thermal Conductors and Insulators . . . 15

1.7 Magnetic and Non-magnetic Materials . . . 17

1.8 Summary . . . 18

2 What are the objects around us made of? - Grade 10 **21**

2.1 Introduction: The atom as the building block of matter . . . 21

2.2 Molecules . . . 21

2.2.1 Representing molecules . . . 21

2.3 Intramolecular and intermolecular forces . . . 25

2.4 The Kinetic Theory of Matter . . . 26

2.5 The Properties of Matter . . . 28

2.6 Summary . . . 31

3 The Atom - Grade 10 **35**

3.1 Models of the Atom . . . 35

3.1.1 The Plum Pudding Model . . . 35

3.1.2 Rutherford's model of the atom . . . 36

3.1.3 The Bohr Model . . . 37
3.2 How big is an atom? . . . 38
3.2.1 How heavy is an atom? . . . 38
3.2.2 How big is an atom? . . . 38
3.3 Atomic structure . . . 38
3.3.1 The Electron . . . 39
3.3.2 The Nucleus . . . 39
3.4 Atomic number and atomic mass number . . . 40
3.5 Isotopes . . . 42
3.5.1 What is an isotope? . . . 42
3.5.2 Relative atomic mass . . . 45
3.6 Energy quantisation and electron configuration . . . 46
3.6.1 The energy of electrons . . . 46
3.6.2 Energy quantisation and line emission spectra . . . 47
3.6.3 Electron configuration . . . 47
3.6.4 Core and valence electrons . . . 51
3.6.5 The importance of understanding electron configuration . . . 51
3.7 Ionisation Energy and the Periodic Table . . . 53
3.7.1 Ions . . . 53
3.7.2 Ionisation Energy . . . 55
3.8 The Arrangement of Atoms in the Periodic Table . . . 56
3.8.1 Groups in the periodic table . . . 56
3.8.2 Periods in the periodic table . . . 58
3.9 Summary . . . 59

4 Atomic Combinations - Grade 11 63
4.1 Why do atoms bond? . . . 63
4.2 Energy and bonding . . . 63
4.3 What happens when atoms bond? . . . 65
4.4 Covalent Bonding . . . 65
4.4.1 The nature of the covalent bond . . . 65
4.5 Lewis notation and molecular structure . . . 69
4.6 Electronegativity . . . 72
4.6.1 Non-polar and polar covalent bonds . . . 73
4.6.2 Polar molecules . . . 73
4.7 Ionic Bonding . . . 74
4.7.1 The nature of the ionic bond . . . 74
4.7.2 The crystal lattice structure of ionic compounds . . . 76
4.7.3 Properties of Ionic Compounds . . . 76
4.8 Metallic bonds . . . 76
4.8.1 The nature of the metallic bond . . . 76
4.8.2 The properties of metals . . . 77

4.9 Writing chemical formulae . . . 78
4.9.1 The formulae of covalent compounds . . . 78
4.9.2 The formulae of ionic compounds . . . 80
4.10 The Shape of Molecules . . . 82
4.10.1 Valence Shell Electron Pair Repulsion (VSEPR) theory . . . 82
4.10.2 Determining the shape of a molecule . . . 82
4.11 Oxidation numbers . . . 85
4.12 Summary . . . 88

5 Intermolecular Forces - Grade 11 91
5.1 Types of Intermolecular Forces . . . 91
5.2 Understanding intermolecular forces . . . 94
5.3 Intermolecular forces in liquids . . . 96
5.4 Summary . . . 97

6 Solutions and solubility - Grade 11 101
6.1 Types of solutions . . . 101
6.2 Forces and solutions . . . 102
6.3 Solubility . . . 103
6.4 Summary . . . 106

7 Atomic Nuclei - Grade 11 107
7.1 Nuclear structure and stability . . . 107
7.2 The Discovery of Radiation . . . 107
7.3 Radioactivity and Types of Radiation . . . 108
7.3.1 Alpha (α) particles and alpha decay . . . 109
7.3.2 Beta (β) particles and beta decay . . . 109
7.3.3 Gamma (γ) rays and gamma decay . . . 110
7.4 Sources of radiation . . . 112
7.4.1 Natural background radiation . . . 112
7.4.2 Man-made sources of radiation . . . 113
7.5 The 'half-life' of an element . . . 113
7.6 The Dangers of Radiation . . . 116
7.7 The Uses of Radiation . . . 117
7.8 Nuclear Fission . . . 118
7.8.1 The Atomic bomb - an abuse of nuclear fission . . . 119
7.8.2 Nuclear power - harnessing energy . . . 120
7.9 Nuclear Fusion . . . 120
7.10 Nucleosynthesis . . . 121
7.10.1 Age of Nucleosynthesis (225 s - 10^3 s) . . . 121
7.10.2 Age of Ions (10^3 s - 10^{13} s) . . . 122
7.10.3 Age of Atoms (10^{13} s - 10^{15} s) . . . 122
7.10.4 Age of Stars and Galaxies (the universe today) . . . 122
7.11 Summary . . . 122

8 Thermal Properties and Ideal Gases - Grade 11 **125**
8.1 A review of the kinetic theory of matter . . . 125
8.2 Boyle's Law: Pressure and volume of an enclosed gas . . . 126
8.3 Charles's Law: Volume and Temperature of an enclosed gas . . . 132
8.4 The relationship between temperature and pressure . . . 136
8.5 The general gas equation . . . 137
8.6 The ideal gas equation . . . 140
8.7 Molar volume of gases . . . 145
8.8 Ideal gases and non-ideal gas behaviour . . . 146
8.9 Summary . . . 147

9 Organic Molecules - Grade 12 **151**
9.1 What is organic chemistry? . . . 151
9.2 Sources of carbon . . . 151
9.3 Unique properties of carbon . . . 152
9.4 Representing organic compounds . . . 152
9.4.1 Molecular formula . . . 152
9.4.2 Structural formula . . . 153
9.4.3 Condensed structural formula . . . 153
9.5 Isomerism in organic compounds . . . 154
9.6 Functional groups . . . 155
9.7 The Hydrocarbons . . . 155
9.7.1 The Alkanes . . . 158
9.7.2 Naming the alkanes . . . 159
9.7.3 Properties of the alkanes . . . 163
9.7.4 Reactions of the alkanes . . . 163
9.7.5 The alkenes . . . 166
9.7.6 Naming the alkenes . . . 166
9.7.7 The properties of the alkenes . . . 169
9.7.8 Reactions of the alkenes . . . 169
9.7.9 The Alkynes . . . 171
9.7.10 Naming the alkynes . . . 171
9.8 The Alcohols . . . 172
9.8.1 Naming the alcohols . . . 173
9.8.2 Physical and chemical properties of the alcohols . . . 175
9.9 Carboxylic Acids . . . 176
9.9.1 Physical Properties . . . 177
9.9.2 Derivatives of carboxylic acids: The esters . . . 178
9.10 The Amino Group . . . 178
9.11 The Carbonyl Group . . . 178
9.12 Summary . . . 179

10 Organic Macromolecules - Grade 12 **185**
10.1 Polymers . . . 185
10.2 How do polymers form? . . . 186
10.2.1 Addition polymerisation . . . 186
10.2.2 Condensation polymerisation . . . 188
10.3 The chemical properties of polymers . . . 190
10.4 Types of polymers . . . 191
10.5 Plastics . . . 191
10.5.1 The uses of plastics . . . 192
10.5.2 Thermoplastics and thermosetting plastics . . . 194
10.5.3 Plastics and the environment . . . 195
10.6 Biological Macromolecules . . . 196
10.6.1 Carbohydrates . . . 197
10.6.2 Proteins . . . 199
10.6.3 Nucleic Acids . . . 202
10.7 Summary . . . 204

III Chemical Change **209**

11 Physical and Chemical Change - Grade 10 **211**
11.1 Physical changes in matter . . . 211
11.2 Chemical Changes in Matter . . . 212
11.2.1 Decomposition reactions . . . 213
11.2.2 Synthesis reactions . . . 214
11.3 Energy changes in chemical reactions . . . 217
11.4 Conservation of atoms and mass in reactions . . . 217
11.5 Law of constant composition . . . 219
11.6 Volume relationships in gases . . . 219
11.7 Summary . . . 220

12 Representing Chemical Change - Grade 10 **223**
12.1 Chemical symbols . . . 223
12.2 Writing chemical formulae . . . 224
12.3 Balancing chemical equations . . . 224
12.3.1 The law of conservation of mass . . . 224
12.3.2 Steps to balance a chemical equation . . . 226
12.4 State symbols and other information . . . 230
12.5 Summary . . . 232

13 Quantitative Aspects of Chemical Change - Grade 11 **233**
13.1 The Mole . . . 233
13.2 Molar Mass . . . 235
13.3 An equation to calculate moles and mass in chemical reactions . . . 237

13.4 Molecules and compounds . . . 239
13.5 The Composition of Substances . . . 242
13.6 Molar Volumes of Gases . . . 246
13.7 Molar concentrations in liquids . . . 247
13.8 Stoichiometric calculations . . . 249
13.9 Summary . . . 252

14 Energy Changes In Chemical Reactions - Grade 11 255
14.1 What causes the energy changes in chemical reactions? . . . 255
14.2 Exothermic and endothermic reactions . . . 255
14.3 The heat of reaction . . . 257
14.4 Examples of endothermic and exothermic reactions . . . 259
14.5 Spontaneous and non-spontaneous reactions . . . 260
14.6 Activation energy and the activated complex . . . 261
14.7 Summary . . . 264

15 Types of Reactions - Grade 11 267
15.1 Acid-base reactions . . . 267
15.1.1 What are acids and bases? . . . 267
15.1.2 Defining acids and bases . . . 267
15.1.3 Conjugate acid-base pairs . . . 269
15.1.4 Acid-base reactions . . . 270
15.1.5 Acid-carbonate reactions . . . 274
15.2 Redox reactions . . . 276
15.2.1 Oxidation and reduction . . . 277
15.2.2 Redox reactions . . . 278
15.3 Addition, substitution and elimination reactions . . . 280
15.3.1 Addition reactions . . . 280
15.3.2 Elimination reactions . . . 281
15.3.3 Substitution reactions . . . 282
15.4 Summary . . . 283

16 Reaction Rates - Grade 12 287
16.1 Introduction . . . 287
16.2 Factors affecting reaction rates . . . 289
16.3 Reaction rates and collision theory . . . 293
16.4 Measuring Rates of Reaction . . . 295
16.5 Mechanism of reaction and catalysis . . . 297
16.6 Chemical equilibrium . . . 300
16.6.1 Open and closed systems . . . 302
16.6.2 Reversible reactions . . . 302
16.6.3 Chemical equilibrium . . . 303
16.7 The equilibrium constant . . . 304

16.7.1 Calculating the equilibrium constant . . . 305
16.7.2 The meaning of k_c values . . . 306
16.8 Le Chatelier's principle . . . 310
16.8.1 The effect of concentration on equilibrium . . . 310
16.8.2 The effect of temperature on equilibrium . . . 310
16.8.3 The effect of pressure on equilibrium . . . 312
16.9 Industrial applications . . . 315
16.10 Summary . . . 316

17 Electrochemical Reactions - Grade 12 319
17.1 Introduction . . . 319
17.2 The Galvanic Cell . . . 320
17.2.1 Half-cell reactions in the Zn-Cu cell . . . 321
17.2.2 Components of the Zn-Cu cell . . . 322
17.2.3 The Galvanic cell . . . 323
17.2.4 Uses and applications of the galvanic cell . . . 324
17.3 The Electrolytic cell . . . 325
17.3.1 The electrolysis of copper sulphate . . . 326
17.3.2 The electrolysis of water . . . 327
17.3.3 A comparison of galvanic and electrolytic cells . . . 328
17.4 Standard Electrode Potentials . . . 328
17.4.1 The different reactivities of metals . . . 329
17.4.2 Equilibrium reactions in half cells . . . 329
17.4.3 Measuring electrode potential . . . 330
17.4.4 The standard hydrogen electrode . . . 330
17.4.5 Standard electrode potentials . . . 333
17.4.6 Combining half cells . . . 337
17.4.7 Uses of standard electrode potential . . . 338
17.5 Balancing redox reactions . . . 342
17.6 Applications of electrochemistry . . . 347
17.6.1 Electroplating . . . 347
17.6.2 The production of chlorine . . . 348
17.6.3 Extraction of aluminium . . . 349
17.7 Summary . . . 349

IV Chemical Systems 353

18 The Water Cycle - Grade 10 355
18.1 Introduction . . . 355
18.2 The importance of water . . . 355
18.3 The movement of water through the water cycle . . . 356
18.4 The microscopic structure of water . . . 359

18.4.1 The polar nature of water . . . 359
18.4.2 Hydrogen bonding in water molecules . . . 359
18.5 The unique properties of water . . . 360
18.6 Water conservation . . . 363
18.7 Summary . . . 366

19 Global Cycles: The Nitrogen Cycle - Grade 10 369
19.1 Introduction . . . 369
19.2 Nitrogen fixation . . . 369
19.3 Nitrification . . . 371
19.4 Denitrification . . . 372
19.5 Human Influences on the Nitrogen Cycle . . . 372
19.6 The industrial fixation of nitrogen . . . 373
19.7 Summary . . . 374

20 The Hydrosphere - Grade 10 377
20.1 Introduction . . . 377
20.2 Interactions of the hydrosphere . . . 377
20.3 Exploring the Hydrosphere . . . 378
20.4 The Importance of the Hydrosphere . . . 379
20.5 Ions in aqueous solution . . . 379
20.5.1 Dissociation in water . . . 380
20.5.2 Ions and water hardness . . . 382
20.5.3 The pH scale . . . 382
20.5.4 Acid rain . . . 384
20.6 Electrolytes, ionisation and conductivity . . . 386
20.6.1 Electrolytes . . . 386
20.6.2 Non-electrolytes . . . 387
20.6.3 Factors that affect the conductivity of water . . . 387
20.7 Precipitation reactions . . . 389
20.8 Testing for common anions in solution . . . 391
20.8.1 Test for a chloride . . . 391
20.8.2 Test for a sulphate . . . 391
20.8.3 Test for a carbonate . . . 392
20.8.4 Test for bromides and iodides . . . 392
20.9 Threats to the Hydrosphere . . . 393
20.10Summary . . . 394

21 The Lithosphere - Grade 11 397
21.1 Introduction . . . 397
21.2 The chemistry of the earth's crust . . . 398
21.3 A brief history of mineral use . . . 399
21.4 Energy resources and their uses . . . 400

21.5 Mining and Mineral Processing: Gold . . . 401
21.5.1 Introduction . . . 401
21.5.2 Mining the Gold . . . 401
21.5.3 Processing the gold ore . . . 401
21.5.4 Characteristics and uses of gold . . . 402
21.5.5 Environmental impacts of gold mining . . . 404
21.6 Mining and mineral processing: Iron . . . 406
21.6.1 Iron mining and iron ore processing . . . 406
21.6.2 Types of iron . . . 407
21.6.3 Iron in South Africa . . . 408
21.7 Mining and mineral processing: Phosphates . . . 409
21.7.1 Mining phosphates . . . 409
21.7.2 Uses of phosphates . . . 409
21.8 Energy resources and their uses: Coal . . . 411
21.8.1 The formation of coal . . . 411
21.8.2 How coal is removed from the ground . . . 411
21.8.3 The uses of coal . . . 412
21.8.4 Coal and the South African economy . . . 412
21.8.5 The environmental impacts of coal mining . . . 413
21.9 Energy resources and their uses: Oil . . . 414
21.9.1 How oil is formed . . . 414
21.9.2 Extracting oil . . . 414
21.9.3 Other oil products . . . 415
21.9.4 The environmental impacts of oil extraction and use . . . 415
21.10 Alternative energy resources . . . 415
21.11 Summary . . . 417

22 The Atmosphere - Grade 11 421
22.1 The composition of the atmosphere . . . 421
22.2 The structure of the atmosphere . . . 422
22.2.1 The troposphere . . . 422
22.2.2 The stratosphere . . . 422
22.2.3 The mesosphere . . . 424
22.2.4 The thermosphere . . . 424
22.3 Greenhouse gases and global warming . . . 426
22.3.1 The heating of the atmosphere . . . 426
22.3.2 The greenhouse gases and global warming . . . 426
22.3.3 The consequences of global warming . . . 429
22.3.4 Taking action to combat global warming . . . 430
22.4 Summary . . . 431

23 The Chemical Industry - Grade 12 **435**

23.1 Introduction 435

23.2 Sasol 435

23.2.1 Sasol today: Technology and production 436

23.2.2 Sasol and the environment 440

23.3 The Chloralkali Industry 442

23.3.1 The Industrial Production of Chlorine and Sodium Hydroxide 442

23.3.2 Soaps and Detergents 446

23.4 The Fertiliser Industry 450

23.4.1 The value of nutrients 450

23.4.2 The Role of fertilisers 450

23.4.3 The Industrial Production of Fertilisers 451

23.4.4 Fertilisers and the Environment: Eutrophication 454

23.5 Electrochemistry and batteries 456

23.5.1 How batteries work 456

23.5.2 Battery capacity and energy 457

23.5.3 Lead-acid batteries 457

23.5.4 The zinc-carbon dry cell 459

23.5.5 Environmental considerations 460

23.6 Summary 461

A GNU Free Documentation License **467**

Part I

Introduction

Part II

Matter and Materials

Chapter 1

Classification of Matter - Grade 10

All the objects that we see in the world around us, are made of **matter**. Matter makes up the air we breathe, the ground we walk on, the food we eat and the animals and plants that live around us. Even our own human bodies are made of matter!

Different objects can be made of different types of matter, or **materials**. For example, a cupboard (an *object*) is made of wood, nails and hinges (the *materials*). The **properties** of the materials will affect the properties of the object. In the example of the cupboard, the strength of the wood and metals make the cupboard strong and durable. In the same way, the raincoats that you wear during bad weather, are made of a material that is waterproof. The electrical wires in your home are made of metal because metals are a type of material that is able to conduct electricity. It is very important to understand the properties of materials, so that we can use them in our homes, in industry and in other applications. In this chapter, we will be looking at different types of materials and their properties.

The diagram below shows one way in which matter can be classified (grouped) according to its different properties. As you read further in this chapter, you will see that there are also other ways of classifying materials, for example according to whether they are good electrical conductors.

Figure 1.1: The classification of matter

1.1 Mixtures

We see mixtures all the time in our everyday lives. A stew, for example, is a mixture of different foods such as meat and vegetables; sea water is a mixture of water, salt and other substances, and air is a mixture of gases such as carbon dioxide, oxygen and nitrogen.

Definition: Mixture
A **mixture** is a combination of more than one substance, where these substances are not bonded to each other.

In a mixture, the substances that make up the mixture:

- *are not in a fixed ratio*

  Imagine, for example, that you have a 250 ml beaker of water. It doesn't matter whether you add 20 g, 40 g, 100 g or any other mass of sand to the water; it will still be called a mixture of sand and water.

- *keep their physical properties*

  In the example we used of the sand and water, neither of these substances has changed in any way when they are mixed together. Even though the sand is in water, it still has the same properties as when it was out of the water.

- *can be separated by mechanical means*

  To separate something by 'mechanical means', means that there is no chemical process involved. In our sand and water example, it is possible to separate the mixture by simply pouring the water through a filter. Something *physical* is done to the mixture, rather than something *chemical*.

Some other examples of mixtures include blood (a mixture of blood cells, platelets and plasma), steel (a mixture of iron and other materials) and the gold that is used to make jewellery. The gold in jewellery is not pure gold but is a mixture of metals. The *carat* of the gold gives an idea of how much gold is in the item.

We can group mixtures further by dividing them into those that are heterogeneous and those that are homogeneous.

1.1.1 Heterogeneous mixtures

A **heterogeneous** mixture does not have a definite composition. Think of a pizza, that is a mixture of cheese, tomato, mushrooms and peppers. Each slice will probably be slightly different from the next because the toppings like the mushrooms and peppers are not evenly distributed. Another example would be granite, a type of rock. Granite is made up of lots of different mineral substances including quartz and feldspar. But these minerals are not spread evenly through the rock and so some parts of the rock may have more quartz than others. Another example is a mixture of oil and water. Although you may add one substance to the other, they will stay separate in the mixture. We say that these heterogeneous mixtures are *non-uniform*, in other words they are not exactly the same throughout.

Definition: Heterogeneous mixture
A heterogeneous mixture is one that is non-uniform, and where the different components of the mixture can be seen.

1.1.2 Homogeneous mixtures

A **homogeneous** mixture has a definite composition, and specific properties. In a homogeneous mixture, the different parts cannot be seen. A solution of salt dissolved in water is an example of a homogeneous mixture. When the salt dissolves, it will spread evenly through the water so that all parts of the solution are the same, and you can no longer see the salt as being separate from the water. Think also of a powdered drink that you mix with water. Provided you give the container a good shake after you have added the powder to the water, the drink will have the same sweet taste for anyone who drinks it, it won't matter whether they take a sip from the top

or from the bottom. The air we breathe is another example of a homogeneous mixture since it is made up of different gases which are in a constant ratio, and which can't be distinguished from each other.

Definition: Homogeneous mixture
A homogeneous mixture is one that is uniform, and where the different components of the mixture cannot be seen.

An **alloy** is a homogeneous mixture of two or more elements, at least one of which is a metal, where the resulting material has metallic properties. Alloys are usually made to improve on the properties of the elements that make them up. Steel for example, is much stronger than iron, which is its main component.

1.1.3 Separating mixtures

Sometimes it is important to be able to separate a mixture. There are lots of different ways to do this. These are some examples:

- *Filtration*

  A piece of filter paper in a funnel can be used to separate a mixture of sand and water.

- *Heating / evaporation*

  Sometimes, heating a solution causes the water to evaporate, leaving the other part of the mixture behind. You can try this using a salt solution.

- *Centrifugation*

  This is a laboratory process which uses the centrifugal force of spinning objects to separate out the heavier substances from a mixture. This process is used to separate the cells and plasma in blood. When the test tubes that hold the blood are spun round in the machine, the heavier cells sink to the bottom of the test tube. Can you think of a reason why it might be important to have a way of separating blood in this way?

- *Dialysis*

  This is an interesting way of separating a mixture because it can be used in some important applications. Dialysis works using a process called *diffusion*. Diffusion takes place when one substance in a mixture moves from an area where it has a high concentration to an area where its concentration is lower. This movement takes place across a semi-permeable membrane. A semi-permeable membrane is a barrier that lets some things move across it, but not others. This process is very important for people whose kidneys are not functioning properly, an illness called *renal failure*.

Interesting Fact

Normally, healthy kidneys remove waste products from the blood. When a person has renal failure, their kidneys cannot do this any more, and this can be life-threatening. Using dialysis, the blood of the patient flows on one side of a semi-permeable membrane. On the other side there will be a fluid that has no waste products but lots of other important substances such as potassium ions (K^+) that the person will need. Waste products from the blood diffuse from where their concentration is high (i.e. in the person's blood) into the 'clean' fluid on the other side of the membrane. The potassium ions will move in the opposite direction from the fluid into the blood. Through this process, waste products are taken out of the blood so that the person stays healthy.

Activity :: Investigation : The separation of a salt solution

Aim:

To demonstrate that a homogeneous salt solution can be separated using physical methods.

Apparatus:

glass beaker, salt, water, retort stand, bunsen burner.

Method:

1. Pour a small amount of water (about 20 ml) into a beaker.
2. Measure a teaspoon of salt and pour this into the water.
3. Stir until the salt dissolves completely. This is now called a *salt solution*. This salt solution is a homogeneous mixture.
4. Place the beaker on a retort stand over a bunsen burner and heat gently. You should increase the heat until the water almost boils.
5. Watch the beaker until all the water has evaporated. What do you see in the beaker?

Results:

The water evaporates from the beaker and tiny grains of salt remain at the bottom.

Conclusion:

The sodium chloride solution, which was a homogeneous mixture of salt and water, has been separated using heating and evaporation.

Activity :: Discussion : Separating mixtures

Work in groups of 3-4

Imagine that you have been given a container which holds a mixture of sand, iron filings (small pieces of iron metal), salt and small stones of different sizes. Is this a homogeneous or a heterogeneous mixture? In your group, discuss how you would go about separating this mixture into the four materials that it contains.

Exercise: Mixtures

1. Which of the following subtances are *mixtures*?
   (a) tap water
   (b) brass (an alloy of copper and zinc)
   (c) concrete
   (d) aluminium
   (e) Coca cola
   (f) distilled water
2. In each of the examples above, say whether the mixture is homogeneous or heterogeneous

1.2 Pure Substances: Elements and Compounds

Any material that is not a mixture, is called a **pure substance**. Pure substances include **elements** and **compounds**. It is much more difficult to break down pure substances into their parts, and complex chemical methods are needed to do this.

1.2.1 Elements

An **element** is a chemical substance that can't be divided or changed into other chemical substances by any ordinary chemical means. The smallest unit of an element is the **atom**.

Definition: Element
An element is a substance that cannot be broken down into other substances through chemical means.

There are 109 known elements. Most of these are natural, but some are man-made. The elements we know are represented in the **Periodic Table of the Elements**, where each element is abbreviated to a **chemical symbol**. Examples of elements are magnesium (Mg), hydrogen (H), oxygen (O) and carbon (C). On the Periodic Table you will notice that some of the abbreviations do not seem to match the elements they represent. The element iron, for example, has the chemical formula Fe. This is because the elements were originally given Latin names. Iron has the abbreviation Fe because its Latin name is 'ferrum'. In the same way, sodium's Latin name is 'natrium' (Na) and gold's is 'aurum' (Au).

1.2.2 Compounds

A **compound** is a chemical substance that forms when two or more elements combine in a fixed ratio. Water (H_2O), for example, is a compound that is made up of two hydrogen atoms for every one oxygen atom. Sodium chloride (NaCl) is a compound made up of one sodium atom for every chlorine atom. An important characteristic of a compound is that it has a **chemical formula**, which describes the ratio in which the atoms of each element in the compound occur.

Definition: Compound
A substance made up of two or more elements that are joined together in a fixed ratio.

Diagram 1.2 might help you to understand the difference between the terms *element*, *mixture* and *compound*. Iron (Fe) and sulfur (S) are two elements. When they are added together, they

Figure 1.2: Understanding the difference between a mixture and a compound

form a *mixture* or iron and sulfur. The iron and sulfur are not joined together. However, if the mixture is heated, a new *compound* is formed, which is called iron sulfide (FeS). In this compound, the iron and sulfur are joined to each other in a ratio of 1:1. In other words, one atom of iron is joined to one atom of sulfur in the compound iron sulfide.

?

Exercise: Elements, mixtures and compounds

1. In the following table, tick whether each of the substances listed is a *mixture* or a *pure substance*. If it is a mixture, also say whether it is a homogeneous or heterogeneous mixture.

Substance	**Mixture or pure**	**Homogeneous or heterogeneous mixture**
fizzy colddrink		
steel		
oxygen		
iron filings		
smoke		
limestone ($CaCO_3$)		

2. In each of the following cases, say whether the substance is an element, a mixture or a compound.
   (a) Cu
   (b) iron and sulfur
   (c) Al
   (d) H_2SO_4
   (e) SO_3

1.3 Giving names and formulae to substances

It is easy to describe elements and mixtures. But how are compounds named? In the example of iron sulfide that was used earlier, which element is named first, and which 'ending' is given to the compound name (in this case, the ending is -ide)?

The following are some guidelines for naming compounds:

1. The compound name will always include the **names of the elements** that are part of it.
   - A compound of **iron** (Fe) and *sulfur* (S) is **iron** *sulf*ide (FeS)
   - A compound of **potassium** (K) and *bromine* (S) is **potassium** *brom*ide (KBr)
   - A compound of **sodium** (Na) and *chlorine* (Cl) is **sodium** *chlor*ide (NaCl)
2. In a compound, the element that is to the left and lower down on the Periodic Table, is used *first* when naming the compound. In the example of NaCl, sodium is a group 1 element on the left hand side of the table, while chlorine is in group 7 on the right of the table. Sodium therefore comes first in the compound name. The same is true for FeS and KBr.
3. The **symbols** of the elements can be used to represent compounds e.g. FeS, NaCl and KBr. These are called **chemical formulae**. In these three examples, the ratio of the elements in each compound is 1:1. So, for FeS, there is one atom of iron for every atom of sulfur in the compound.
4. A compound may contain **compound ions**. Some of the more common compound ions and their names are shown below.

Name of compound ion	formula
Carbonate	$CO_3{}^{2-}$
sulphate	$SO_4{}^{2-}$
Hydroxide	OH^-
Ammonium	$NH_4{}^+$
Nitrate	$NO_3{}^-$
Hydrogen carbonate	$HCO_3{}^-$
Phosphate	$PO_4{}^{3-}$
Chlorate	$ClO_3{}^-$
Cyanide	CN^-
Chromate	$CrO_4{}^{2-}$
Permanganate	$MnO_4{}^-$

5. When there are only two elements in the compound, the compound is often given a **suffix** (ending) of -ide. You would have seen this in some of the examples we have used so far. When a non-metal is combined with oxygen to form a negative ion (anion) which then combines with a positive ion (cation) from hydrogen or a metal, then the suffix of the name will be ...ate or ...ite. NO_3^- for example, is a negative ion, which may combine with a cation such as hydrogen (HNO_3) or a metal like potassium (KNO_3). The NO_3^- anion has the name nitr**ate**. SO_3 in a formula is sulph**ite**, e.g. sodium sulphite (Na_2SO_3). SO_4 is sulph**ate** and PO_4 is phosph**ate**.
6. **Prefixes** can be used to describe the ratio of the elements that are in the compound. You should know the following prefixes: 'mono' (one), 'di' (two) and 'tri' (three).
   - CO (carbon monoxide) - There is one atom of oxygen for every one atom of carbon
   - NO_2 (nitrogen dioxide) - There are two atoms of oxygen for every one atom of nitrogen
   - SO_3 (sulfur trioxide) - There are three atoms of oxygen for every one atom of sulfur

Important:
When numbers are written as 'subscripts' in compounds (i.e. they are written below the element symbol), this tells us how many atoms of that element there are in relation to other elements in the compound. For example in nitrogen dioxide (NO_2) there are two oxygen atoms for every one atom of nitrogen. In sulfur trioxide (SO_3), there are three oxygen atoms for every one atom of sulfur in the compound. Later, when we start looking at chemical equations, you will notice that sometimes there are numbers *before* the compound name. For example, $2H_2O$ means that there are two molecules of water, and that in each molecule there are two hydrogen atoms for every one oxygen atom.

Exercise: Naming compounds

1. The formula for calcium carbonate is $CaCO_3$.
   (a) Is calcium carbonate a mixture or a compound? Give a reason for your answer.
   (b) What is the ratio of Ca:C:O atoms in the formula?
2. Give the name of each of the following substances.
   (a) KBr
   (b) HCl
   (c) $KMnO_4$
   (d) NO_2
   (e) NH_4OH
   (f) Na_2SO_4
3. Give the chemical formula for each of the following compounds.
   (a) potassium nitrate
   (b) sodium iodide
   (c) barium sulphate
   (d) nitrogen dioxide
   (e) sodium monosulphate
4. Refer to the diagram below, showing sodium chloride and water, and then answer the questions that follow.

   (a) What is the chemical formula for water?
   (b) What is the chemical formula for sodium chloride?
   (c) Label the water and sodium chloride in the diagram.
   (d) Which of the following statements most accurately describes the picture?
      i. The picture shows a mixture of an element and a compound
      ii. The picture shows a mixture of two compounds
      iii. The picture shows two compounds that have been chemically bonded to each other
5. What is the formula of this molecule?

```
      H     H
      |     |
H — C — C — O — H
      |     |
      H     H
```

   A C_6H_2O
   B C_2H_6O
   C 2C6HO
   D $_2CH_6O$

1.4 Metals, Semi-metals and Non-metals

The elements in the Periodic Table can also be divided according to whether they are **metals**, **semi-metals** or **non-metals**. On the right hand side of the Periodic Table is a dark 'zigzag' line. This line separates all the elements that are metals from those that are non-metals. Metals are found on the left of the line, and non-metals are those on the right. Metals, semi-metals and non-metals all have their own specific properties.

1.4.1 Metals

Examples of metals include copper (Cu), zinc (Zn), gold (Au) and silver (Ag). On the Periodic Table, the metals are on the left of the zig-zag line. There are a large number of elements that are metals. The following are some of the properties of metals:

- *Thermal conductors*

  Metals are good conductors of heat and are therefore used in cooking utensils such as pots and pans.

- *Electrical conductors*

  Metals are good conductors of electricity, and are therefore used in electrical conducting wires.

- *Shiny metallic lustre*

  Metals have a characteristic shiny appearance and are often used to make jewellery.

- *Malleable*

  This means that they can be bent into shape without breaking.

- *Ductile*

  Metals can stretched into thin wires such as copper, which can then be used to conduct electricity.

- *Melting point*

  Metals usually have a high melting point and can therefore be used to make cooking pots and other equipment that needs to become very hot, without being damaged.

You can see how the properties of metals make them very useful in certain applications.

Activity :: Group Work : Looking at metals

1. Collect a number of metal items from your home or school. Some examples are listed below:
   - hammer
   - electrical wiring
   - cooking pots
   - jewellery
   - burglar bars
   - coins
2. In groups of 3-4, combine your collection of metal objects.
3. What is the function of each of these objects?
4. Discuss why you think metal was used to make each object. You should consider the properties of metals when you answer this question.

1.4.2 Non-metals

In contrast to metals, non-metals are poor thermal conductors, good electrical insulators (meaning that they do *not* conduct electrical charge) and are neither malleable nor ductile. The non-metals are found on the right hand side of the Periodic Table, and include elements such as sulfur (S), phosphorus (P), nitrogen (N) and oxygen (O).

1.4.3 Semi-metals

Semi-metals have mostly non-metallic properties. One of their distinguishing characteristics is that their conductivity increases as their temperature increases. This is the opposite of what happens in metals. The semi-metals include elements such as silicon (Si) and germanium (Ge). Notice where these elements are positioned in the Periodic Table.

1.5 Electrical conductors, semi-conductors and insulators

An **electrical conductor** is a substance that allows an electrical current to pass through it. Many electrical conductors are metals, but non-metals can also be good conductors. *Copper* is one of the best electrical conductors, and this is why it is used to make conducting wire. In reality, *silver* actually has an even higher electrical conductivity than copper, but because silver is so expensive, it is not practical to use it for electrical wiring because such large amounts are needed. In the overhead power lines that we see above us, *aluminium* is used. The aluminium usually surrounds a steel core which adds tensile strength to the metal so that it doesn't break when it is stretched across distances. Occasionally gold is used to make wire, not because it is a particularly good conductor, but because it is very resistant to surface corrosion. *Corrosion* is when a material starts to deteriorate at the surface because of its reactions with the surroundings, for example oxygen and water in the air.

An **insulator** is a non-conducting material that does not carry any charge. Examples of insulators would be plastic and wood. Do you understand now why electrical wires are normally covered with plastic insulation? **Semi-conductors** behave like insulators when they are cold, and like conductors when they are hot. The elements silicon and germanium are examples of semi-conductors.

Definition: Conductors and insulators
A conductor allows the easy movement or flow of something such as heat or electrical charge through it. Insulators are the opposite to conductors because they *inhibit* or reduce the flow of heat, electrical charge, sound etc through them.

Activity :: Experiment : Electrical conductivity

Aim:
To investigate the electrical conductivity of a number of substances

Apparatus:

- two or three cells
- light bulb
- crocodile clips
- wire leads
- a selection of test substances (e.g. a piece of plastic, aluminium can, metal pencil sharpener, metal magnet, wood, chalk).

Method:

1. Set up the circuit as shown above, so that the test substance is held between the two crocodile clips. The wire leads should be connected to the cells and the light bulb should also be connected into the circuit.
2. Place the test substances one by one between the crocodile clips and see what happens to the light bulb.

Results:

Record your results in the table below:

Test substance	Metal/non-metal	Does bulb glow?	Conductor or insulator

Conclusions:

In the substances that were tested, the metals were able to conduct electricity and the non-metals were not. Metals are good electrical conductors and non-metals are not.

1.6 Thermal Conductors and Insulators

A **thermal conductor** is a material that allows energy in the form of heat, to be transferred within the material, without any movement of the material itself. An easy way to understand this concept is through a simple demonstration.

Activity :: Demonstration : Thermal conductivity
Aim:
To demonstrate the ability of different substances to conduct heat.
Apparatus:

You will need two cups (made from the same material e.g. plastic); a metal spoon and a plastic spoon.

Method:

- Pour boiling water into the two cups so that they are about half full.
- At the same time, place a metal spoon into one cup and a plastic spoon in the other.
- Note which spoon heats up more quickly

Results:

The metal spoon heats up more quickly than the plastic spoon. In other words, the metal conducts heat well, but the plastic does not.

Conclusion:

Metal is a good thermal conductor, while plastic is a poor thermal conductor. This explains why cooking pots are metal, but their handles are often plastic or wooden. The pot itself must be metal so that heat from the cooking surface can heat up the pot to cook the food inside it, but the handle is made from a poor thermal conductor so that the heat does not burn the hand of the person who is cooking.

An **insulator** is a material that does not allow a transfer of electricity or energy. Materials that are poor thermal conductors can also be described as being good insulators.

Interesting Fact

Water is a better thermal conductor than air and conducts heat away from the body about 20 times more efficiently than air. A person who is not wearing a wetsuit, will lose heat very quickly to the water around them and can be vulnerable to hypothermia. Wetsuits help to preserve body heat by trapping a layer of water against the skin. This water is then warmed by body heat and acts as an insulator. Wetsuits are made out of closed-cell, foam neoprene. Neoprene is a synthetic rubber that contains small bubbles of nitrogen gas when made for use as wetsuit material. Nitrogen gas has very low thermal conductivity, so it does not allow heat from the body (or the water trapped between the body and the wetsuit) to be lost to the water outside of the wetsuit. In this way a person in a wetsuit is able to keep their body temperature much higher than they would otherwise.

Activity :: Investigation : A closer look at thermal conductivity

Look at the table below, which shows the thermal conductivity of a number of different materials, and then answer the questions that follow. The higher the number in the second column, the better the material is at conducting heat (i.e. it is a good thermal conductor). Remember that a material that conducts heat efficiently, will also lose heat more quickly than an insulating material.

Material	Thermal Conductivity (W/m/K)
Silver	429
Stainless steel	16
Standard glass	1.05
Concrete	0.9 - 2
Red brick	0.69
Water	0.58
Snow	0.5 - 0.25
Wood	0.04 - 0.12
Polystyrene	0.03
Air	0.024

Use this information to answer the following questions:

1. Name two materials that are good thermal conductors.
2. Name two materials that are good insulators.
3. Explain why:
   (a) cooler boxes are often made of polystyrene
   (b) homes that are made from wood need less internal heating during the winter months.
   (c) igloos (homes made from snow) are so good at maintaining warm temperatures, even in freezing conditions.

Interesting Fact

It is a known fact that well-insulated buildings need less energy for heating than do buildings that have no insulation. Two building materials that are being used more and more worldwide, are **mineral wool** and **polystyrene**. Mineral wool is a good insulator because it holds air still in the matrix of the wool so that heat is not lost. Since air is a poor conductor and a good insulator, this helps to keep energy within the building. Polystyrene is also a good insulator and is able to keep cool things cool and hot things hot! It has the added advantage of being resistant to moisture, mould and mildew.

Remember that concepts such as conductivity and insulation are not only relevant in the building, industrial and home environments. Think for example of the layer of blubber or fat that we find in animals. In very cold environments, fat and blubber not only provide protection, but also act as an insulator to help the animal to keep its body temperature at the right level. This is known as *thermoregulation*.

1.7 Magnetic and Non-magnetic Materials

We have now looked at a number of ways in which matter can be grouped, such as into metals, semi-metals and non-metals; electrical conductors and insulators, and thermal conductors and insulators. One way in which we can further group metals, is to divide them into those that are **magnetic** and those that are **non-magnetic.**

Definition: Magnetism
Magnetism is one of the phenomena by which materials exert attractive or repulsive forces on other materials.

A metal is said to be **ferromagnetic** if it can be magnetised (i.e. made into a magnet). If you hold a magnet very close to a metal object, it may happen that its own electrical field will be induced and the object becomes magnetic. Some metals keep their magnetism for longer than others. Look at iron and steel for example. Iron loses its magnetism quite quickly if it is taken away from the magnet. Steel on the other hand will stay magnetic for a longer time. Steel is often used to make permanent magnets that can be used for a variety of purposes.

Magnets are used to sort the metals in a scrap yard, in compasses to find direction, in the magnetic strips of video tapes and ATM cards where information must be stored, in computers and TV's, as well as in generators and electric motors.

Activity :: Investigation : Magnetism

You can test whether an object is magnetic or not by holding another magnet close to it. If the object is attracted to the magnet, then it too is magnetic.

Find some objects in your classroom or your home and test whether they are magnetic or not. Then complete the table below:

Object	**Magnetic or non-magnetic**

Activity :: Group Discussion : Properties of materials

In groups of 4-5, discuss how our knowledge of the properties of materials has allowed society to:

- develop advanced computer technology
- provide homes with electricity
- find ways to conserve energy

1.8 Summary

- All the objects and substances that we see in the world are made of **matter**.
- This matter can be classified according to whether it is a **mixture** or a **pure substance**.
- A **mixture** is a combination of one or more substances that are not chemically bonded to each other. Examples of mixtures are air (a mixture of different gases) and blood (a mixture of cells, platelets and plasma).
- The main **characteristics** of mixtures are that the substances that make them up are not in a fixed ratio, they keep their individual properties and they can be separated from each other using mechanical means.

- A **heterogeneous mixture** is non-uniform and the different parts of the mixture can be seen. An example would be a mixture of sand and salt.
- A **homogeneous mixture** is uniform, and the different components of the mixture can't be seen. An example would be a salt solution. A salt solution is a mixture of salt and water. The salt dissolves in the water, meaning that you can't see the individual salt particles. They are interspersed between the water molecules. Another example is a metal **alloy** such as steel.
- Mixtures can be **separated** using a number of methods such as filtration, heating, evaporation, centrifugation and dialysis.
- Pure substances can be further divided into **elements** and **compounds**.
- An **element** is a substance that can't be broken down into simpler substances through chemical means.
- All the elements are recorded in the **Periodic Table of the Elements**. Each element has its own chemical symbol. Examples are iron (Fe), sulfur (S), calcium (Ca), magnesium (Mg) and fluorine (F).
- A **compound** is a substance that is made up of two or more elements that are chemically bonded to each other in a fixed ratio. Examples of compounds are sodium chloride (NaCl), iron sulfide (FeS), calcium carbonate ($CaCO_3$) and water (H_2O).
- When **naming compounds** and writing their **chemical formula**, it is important to know the elements that are in the compound, how many atoms of each of these elements will combine in the compound and where the elements are in the Periodic Table. A number of rules can then be followed to name the compound.
- Another way of classifying matter is into **metals** (e.g. iron, gold, copper), **semi-metals** (e.g. silicon and germanium) and **non-metals** (e.g. sulfur, phosphorus and nitrogen).
- **Metals** are good electrical and thermal conductors, they have a shiny lustre, they are malleable and ductile, and they have a high melting point. These properties make metals very useful in electrical wires, cooking utensils, jewellery and many other applications.
- A further way of classifying matter is into **electrical conductors**, **semi-conductors** and **insulators**.
- An **electrical conductor** allows an electrical current to pass through it. Most metals are good electrical conductors.
- An **electrical insulator** is not able to carry an electrical current. Examples are plastic, wood, cotton material and ceramic.
- Materials may also be classified as **thermal conductors** or **thermal insulators** depending on whether or not they are able to conduct heat.
- Materials may also be either **magnetic** or **non-magnetic**.

?

Exercise: Summary

1. For each of the following **multiple choice** questions, choose *one* correct answer from the list provided.

   A Which of the following can be classified as a mixture:

      i. sugar
      ii. table salt
      iii. air
      iv. Iron

   B An element can be defined as:

i. A substance that cannot be separated into two or more substances by ordinary chemical (or physical) means
ii. A substance with constant composition
iii. A substance that contains two or more substances, in definite proportion by weight
iv. A uniform substance

2. Classify each of the following substances as an *element*, a *compound*, a *solution*(homogeneous mixture), or a *heterogeneous mixture*: salt, pure water, soil, salt water, pure air, carbon dioxide, gold and bronze
3. Look at the table below. In the first column (A) is a list of substances. In the second column (B) is a description of the group that each of these substances belongs in. Match up the *substance* in Column A with the *description* in Column B.

Column A	**Column B**
iron	a compound containing 2 elements
H_2S	a heterogeneous mixture
sugar solution	a metal alloy
sand and stones	an element
steel	a homogeneous mixture

4. You are given a test tube that contains a mixture of iron filings and sulfur. You are asked to weigh the amount of iron in the sample.
   a Suggest one method that you could use to separate the iron filings from the sulfur.
   b What property of metals allows you to do this?
5. Given the following descriptions, write the chemical formula for each of the following substances:
   a silver metal
   b a compound that contains only potassium and bromine
   c a gas that contains the elements carbon and oxygen in a ratio of 1:2
6. Give the names of each of the following compounds:
   a NaBr
   b $BaSO_4$
   c SO_2
7. For each of the following materials, say what properties of the material make it important in carrying out its particular function.
   a **tar** on roads
   b **iron** burglar bars
   c **plastic** furniture
   d **metal** jewellery
   e **clay** for building
   f **cotton** clothing

Chapter 2

What are the objects around us made of? - Grade 10

2.1 Introduction: The atom as the building block of matter

We have now seen that different materials have different properties. Some materials are metals and some are non-metals; some are electrical or thermal conductors, while others are not. Depending on the properties of these materials, they can be used in lots of useful applications. But what is it exactly that makes up these materials? In other words, if we were to break down a material into the parts that make it up, what would we find? And how is it that a material's microscopic structure is able to give it all these different properties?

The answer lies in the smallest building block of matter: the **atom**. It is the *type* of atoms, and the way in which they are *arranged* in a material, that affects the properties of that substance.

It is not often that substances are found in atomic form. Normally, atoms are bonded to other atoms to form **compounds** or **molecules**. It is only in the *noble gases* (e.g. helium, neon and argon) that atoms are found individually and are not bonded to other atoms. We will look at the reasons for this in a later chapter.

2.2 Molecules

Definition: Molecule
A molecule is a group of two or more atoms that are attracted to each other by relatively strong forces or bonds

Almost everything around us is made up of molecules. *Water* is made up of molecules, each of which has two hydrogen atoms joined to one oxygen atom. *Oxygen* is a molecule that is made up of two oxygen atoms that are joined to one another. Even the food that we eat is made up of molecules that contain atoms of elements such as carbon, hydrogen and oxygen that are joined to one another in different ways. All of these are known as **small molecules** because there are only a few atoms in each molecule. **Giant molecules** are those where there may be millions of atoms per molecule. Examples of giant molecules are *diamonds*, which are made up of millions of carbon atoms bonded to each other, and *metals*, which are made up of millions of metal atoms bonded to each other.

2.2.1 Representing molecules

The structure of a molecule can be shown in many different ways. Sometimes it is easiest to show what a molecule looks like by using different types of **diagrams**, but at other times, we may decide to simply represent a molecule using its **chemical formula** or its written name.

1. **Using formulae to show the structure of a molecule**

   A **chemical formula** is an abbreviated (shortened) way of describing a molecule, or some other chemical substance. In chapter 1, we saw how chemical compounds can be represented using element symbols from the Periodic Table. A chemical formula can also tell us the *number* of atoms of each element that are in a molecule, and their *ratio* in that molecule.

   For example, the chemical formula for a molecule of carbon dioxide is:

   $$CO_2$$

   The formula above is called the **molecular formula** of that compound. The formula tells us that in one molecule of carbon dioxide, there is one atom of carbon and two atoms of oxygen. The ratio of carbon atoms to oxygen atoms is 1:2.

Definition: Molecular formula
A concise way of expressing information about the atoms that make up a particular chemical compound. The molecular formula gives the exact number of each type of atom in the molecule.

   A molecule of glucose has the molecular formula:

   $$C_6H_{12}O_6$$

   In each glucose molecule, there are six carbon atoms, twelve hydrogen atoms and six oxygen atoms. The ratio of carbon:hydrogen:oxygen is 6:12:6. We can simplify this ratio to write 1:2:1, or if we were to use the element symbols, the formula would be written as CH_2O. This is called the **empirical formula** of the molecule.

Definition: Empirical formula
This is a way of expressing the *relative* number of each type of atom in a chemical compound. In most cases, the empirical formula does not show the exact number of atoms, but rather the simplest *ratio* of the atoms in the compound.

   The empirical formula is useful when we want to write the formula for a *giant molecule*. Since giant molecules may consist of millions of atoms, it is impossible to say exactly how many atoms are in each molecule. It makes sense then to represent these molecules using their empirical formula. So, in the case of a metal such as copper, we would simply write Cu, or if we were to represent a molecule of sodium chloride, we would simply write NaCl.

   Chemical formulae therefore tell us something about the *types* of atoms that are in a molecule and the *ratio* in which these atoms occur in the molecule, but they don't give us any idea of what the molecule actually looks like, in other words its *shape*. Another useful way of representing molecules is to use diagrams.

   Another type of formula that can be used to describe a molecule is its **structural formula**. A structural formula uses a graphical representation to show a molecule's structure (figure 2.1).

2. **Using diagrams to show the structure of a molecule**

   Diagrams of molecules are very useful because they give us an idea of the *space* that is occupied by the molecule, and they also help us to picture how the atoms are arranged in the molecule. There are two types of diagrams that are commonly used:

(a) $\mathbf{C_4H_{10}}$ (b) $\mathbf{C_2H_5}$

Figure 2.1: Diagram showing (a) the molecular, (b) the empirical and (c) the structural formula of isobutane

- *Ball and stick models*
  This is a 3-dimensional molecular model that uses 'balls' to represent atoms and 'sticks' to represent the bonds between them. The centres of the atoms (the balls) are connected by straight lines which represent the bonds between them. A simplified example is shown in figure 2.2.

Figure 2.2: A ball and stick model of a water molecule

- *Space-filling model*
  This is also a 3-dimensional molecular model. The atoms are represented by multi-coloured spheres. Space-filling models of water and ammonia are shown in figures 2.3 and 2.4.

Figures 2.3 and 2.4 are some examples of **simple molecules** that are represented in different ways.

Figure 2.3: A space-filling model and structural formula of a water molecule. Each molecule is made up of two hydrogen atoms that are attached to one oxygen atom. This is a simple molecule.

Figure 2.5 shows the bonds between the carbon atoms in diamond, which is a **giant molecule**. Each carbon atom is joined to four others, and this pattern repeats itself until a complex *lattice* structure is formed. Each black ball in the diagram represents a carbon atom, and each line represents the bond between two carbon atoms.

Interesting Fact

Diamonds are most often thought of in terms of their use in the jewellery industry. However, about 80% of mined diamonds are unsuitable for use as gemstones and are therefore used in industry because of their strength and hardness. These

Figure 2.4: A space-filling model and structural formula of a molecule of ammonia. Each molecule is made up of one nitrogen atom and three hydrogen atoms. This is a simple molecule.

Figure 2.5: Diagrams showing the microscopic structure of diamond. The diagram on the left shows part of a diamond lattice, made up of numerous carbon atoms. The diagram on the right shows how each carbon atom in the lattice is joined to four others. This forms the basis of the lattice structure. Diamond is a giant molecule.

properties of diamonds are due to the strong covalent bonds betwene the carbon atoms in diamond. The most common uses for diamonds in industry are in cutting, drilling, grinding, and polishing.

Exercise: Atoms and molecules

1. In each of the following, say whether the chemical substance is made up of single atoms, simple molecules or giant molecules.
   (a) ammonia gas (NH_3)
   (b) zinc metal (Zn)
   (c) graphite (C)
   (d) nitric acid (HNO_3)
   (e) neon gas (Ne_2)
2. Refer to the diagram below and then answer the questions that follow:

O — C — O

(a) Identify the molecule.
(b) Write the molecular formula for the molecule.
(c) Is the molecule a simple or giant molecule?

3. Represent each of the following molecules using its *chemical formula*, *structural formula* and *ball and stick model*.
   (a) H_2
   (b) NH_3
   (c) sulfur dioxide

2.3 Intramolecular and intermolecular forces

When atoms join to form molecules, they are held together by **chemical bonds**. The type of bond, and the strength of the bond, depends on the atoms that are involved. These bonds are called **intramolecular forces** because they are bonding forces *inside* a molecule ('intra' means 'within' or 'inside'). Sometimes we simply call these intramolecular forces chemical bonds.

Definition: Intramolecular force
The force between the atoms of a molecule, which holds them together.

Examples of the types of chemical bonds that can exist between atoms inside a molecule are shown below. These will be looked at in more detail in chapter 4.

- *Covalent bond*

  Covalent bonds exist between non-metal atoms e.g. There are covalent bonds between the carbon and oxygen atoms in a molecule of carbon dioxide.

- *Ionic bond*

  Ionic bonds occur between non-metal and metal atoms e.g. There are ionic bonds between the sodium and chlorine atoms in a molecule of sodium chloride.

- *Metallic bond*

  Metallic bonds join metal atoms e.g. There are metallic bonds between copper atoms in a piece of copper metal.

Intermolecular forces are those bonds that hold *molecules* together. A glass of water for example, contains many molecules of water. These molecules are held together by intermolecular forces. The strength of the intermolecular forces is important because they affect properties such as *melting point* and *boiling point*. For example, the stronger the intermolecular forces, the higher the melting point and boiling point for that substance. The strength of the intermolecular forces increases as the size of the molecule increases.

Definition: Intermolecular force
A force between molecules, which holds them together.

Diagram 2.6 may help you to understand the difference between intramolecular forces and intermolecular forces.

Figure 2.6: Two representations showing the intermolecular and intramolecular forces in water: space-filling model and structural formula.

It should be clearer now that there are two types of forces that hold matter together. In the case of water, there are intramolecular forces that hold the two hydrogen atoms to the oxygen atom *in each molecule of water*. There are also intramolecular forces *between each of these water molecules*. As mentioned earlier, these forces are very important because they affect many of the *properties of matter* such as boiling point, melting point and a number of other properties. Before we go on to look at some of these examples, it is important that we first take a look at the **Kinetic Theory of Matter**.

Exercise: Intramolecular and intermolecular forces

1. Using ammonia gas as an example...
   (a) Explain what is meant by an *intramolecular* force or *chemical bond*.
   (b) Explain what is meant by an *intermolecular* force.
2. Draw a diagram showing three molecules of carbon dioxide. On the diagram, show where the intramolecular and intermolecular forces are.
3. Why is it important to understand the types of forces that exist between atoms and between molecules? Try to use some practical examples in your answer.

2.4 The Kinetic Theory of Matter

The **kinetic theory of matter** is used to explain why matter exists in different *phases* (i.e. solid, liquid and gas), and how matter can change from one phase to the next. The kinetic theory of matter also helps us to understand other properties of matter. It is important to realise that what we will go on to describe is only a *theory*. It cannot be proved beyond doubt, but the fact that it helps us to explain our observations of changes in phase, and other properties of matter, suggests that it probably is more than just a theory.

Broadly, the Kinetic Theory of Matter says that:

- Matter is made up of **particles** that are constantly moving.
- All particles have **energy**, but the energy varies depending on whether the substance is a solid, liquid or gas. Solid particles have the least energy and gas particles have the most amount of energy.
- The **temperature** of a substance is a measure of the *average kinetic energy* of the particles.
- A change in **phase** may occur when the energy of the particles is changed.
- There are **spaces** between the particles of matter.
- There are **attractive forces** between particles and these become stronger as the particles move closer together. These attractive forces will either be intramolecular forces (if the particles are atoms) or intermolecular forces (if the particles are molecules). When the particles are extremely close, repulsive forces start to act.

Table 2.1 summarises the characteristics of the particles that are in each phase of matter.

Table 2.1: Table summarising the general features of solids, liquids and gases.

Property of matter	**Gas**	**Liquid**	**Gas**
Particles	Atoms or molecules	Atoms or molecules	Atoms or molecules
Energy and movement of particles	Particles have high energy and are constantly moving	Particles have less energy than in the gas phase	Low energy - particles vibrate around a fixed point
Spaces between particles	Large spaces because of high energy	Smaller spaces than in gases	Very little space between particles. Particles are tightly packed together
Attractive forces between particles	Weak forces because of the large distance between particles	Stronger forces than in gas. Liquids can be poured.	Very strong forces. Solids have a fixed volume.
Changes in phase	In general a gas becomes a liquid or solid when it is cooled. Particles have less energy and therefore move closer together so that the attractive forces become stronger, and the gas becomes a liquid or a solid	A liquid becomes a gas if its temperature is increased. It becomes a solid if its temperature decreases.	Solids become liquids or gases if their temperature is increased.

Let's look at an example that involves the three phases of water: ice (solid), water (liquid) and water vapour (gas).

Figure 2.7: The three phases of matter

In a solid (e.g. ice), the water molecules have very little energy and can't move away from each other. The molecules are held close together in a regular pattern called a *lattice*. If the ice is

heated, the energy of the molecules increases. This means that some of the water molecules are able to overcome the intermolecular forces that are holding them together, and the molecules move further apart to form *liquid water*. This is why liquid water is able to flow, because the molecules are more free to move than they were in the solid lattice. If the molecules are heated further, the liquid water will become water vapour, which is a gas. Gas particles have lots of energy and are far away from each other. That is why it is difficult to keep a gas in a specific area! The attractive forces between the particles are very weak and they are only loosely held together. Figure 2.8 shows the changes in phase that may occur in matter, and the names that describe these processes.

Figure 2.8: Changes in phase

2.5 The Properties of Matter

Let us now look at what we have learned about chemical bonds, intermolecular forces and the kinetic theory of matter, and see whether this can help us to understand some of the macroscopic properties of materials.

1. **Melting point**

Definition: Melting point
The temperature at which a *solid* changes its phase or state to become a *liquid*. The reverse process (change in phase from liquid to solid) is called **freezing**.

In order for a solid to melt, the energy of the particles must increase enough to overcome the bonds that are holding the particles together. It makes sense then that a solid which is held together by strong bonds will have a *higher* melting point than one where the bonds are weak, because more energy (heat) is needed to break the bonds. In the examples we have looked at, metals, ionic solids and some atomic lattices (e.g. diamond) have high melting points, whereas the melting points for molecular solids and other atomic lattices (e.g. graphite) are much lower. Generally, the intermolecular forces between molecular solids are *weaker* than those between ionic and metallic solids.

2. **Boiling point**

Definition: Boiling point
The temperature at which a *liquid* changes its phase to become a *gas*.

When the temperature of a liquid increases, the average kinetic energy of the particles also increases, and they are able to overcome the bonding forces that are holding them in the liquid. When boiling point is reached, *evaporation* takes place and some particles in the liquid become a gas. In other words, the energy of the particles is too great for them to be held in a liquid anymore. The stronger the bonds within a liquid, the higher the boiling point needs to be in order to break these bonds. Metallic and ionic compounds have high boiling points while the boiling point for molecular liquids is lower.

The data in table 2.2 below may help you to understand some of the concepts we have explained. Not all of the substances in the table are solids at room temperature, so for now, let's just focus on the *boiling points* for each of these substances. Of the substances listed, ethanol has the weakest intermolecular forces, and sodium chloride and mercury have the strongest. What do you notice?

Substance	**Melting point (^0C)**	**Boiling point (^0C)**
Ethanol (C_2H_6O)	-114,3	78,4
Water	0	100
Mercury	-38,83	356,73
Sodium chloride	801	1465

Table 2.2: The melting and boiling points for a number of substances

You will have seen that substances such as ethanol, with relatively weak intermolecular forces, have the lowest boiling point, while substances with stronger intermolecular forces such as sodium chloride and mercury, must be heated much more if the particles are to have enough energy to overcome the forces that are holding them together in the liquid or solid phase.

?

Exercise: Forces and boiling point

The table below gives the molecular formula and the boiling point for a number of organic compounds called *alkanes*. Refer to the table and then answer the questions that follow.

Organic compound	**Molecular formula**	**Boiling point (^{0}C)**
Methane	CH_2	-161.6
Ethane	C_2H_6	-88.6
Propane	C_3H_8	-45
Butane	C_4H_{10}	-0.5
Pentane	C_5H_{12}	36.1
Hexane	C_6H_{14}	69
Heptane	C_7H_{16}	98.42
Octane	C_8H_{18}	125.52

Data from: http://www.wikipedia.com

(a) Draw a graph to show the relationship between the number of carbon atoms in each alkane, and its boiling point (Number of carbon atoms will go on the x-axis and boiling point on the y-axis).

(b) Describe what you see.

(c) Suggest a reason for what you have observed.

(d) Why was it enough for us to use 'number of carbon atoms' as a measure of the molecular weight of the molecules?

---

3. **Density and viscosity**

Density is a measure of the mass of a substance per unit volume. The density of a solid is generally higher than that of a liquid because the particles are hold much more closely

together and therefore there are more particles packed together in a particular volume. In other words, there is a greater mass of the substance in a particular volume. In general, density increases as the strength of the intermolecular forces increases. **Viscosity** is a measure of how resistant a liquid is to changing its form. Viscosity is also sometimes described as the 'thickness' of a fluid. Think for example of syrup and how slowly it pours from one container into another. Now compare this to how easy it is to pour water. The viscosity of syrup is greater than the viscosity of water. Once again, the stronger the intermolecular forces in the liquid, the greater its viscosity.

It should be clear now that we can explain a lot of the **macroscopic properties** of matter (i.e. the characteristics we can see or observe) by understanding their **microscopic structure** and the way in which the atoms and molecules that make up matter are held together.

Activity :: Investigation : Determining the density of liquids:

Density is a very important property because it helps us to identify different materials. Every material, depending on the elements that make it up, and the arrangement of its atoms, will have a different density.

The equation for density is:

$$\text{Density} = \text{Mass}/\text{Volume}$$

Discussion questions:

To calculate the density of liquids and solids, we need to be able to first determine their mass and volume. As a group, think about the following questions:

- How would you determine the mass of a liquid?
- How would you determine the volume of an irregular solid?

Apparatus:

Laboratory mass balance, 10 ml and 100 ml graduated cylinders, thread, distilled water, two different liquids.

Method:

Determine the density of the distilled water and two liquids as follows:

1. Measure and record the mass of a 10 ml graduated cyclinder.
2. Pour an amount of distilled water into the cylinder.
3. Measure and record the combined mass of the water and cylinder.
4. Record the volume of distilled water in the cylinder
5. Empty, clean and dry the graduated cylinder.
6. Repeat the above steps for the other two liquids you have.
7. Complete the table below.

Liquid	**Mass (g)**	**Volume (ml)**	**Density (g/ml)**
Distilled water			
Liquid 1			
Liquid 2			

Activity :: Investigation : Determining the density of irregular solids:

Apparatus:

Use the same materials and equpiment as before (for the liquids). Also find a number of solids that have an irregular shape.

Method:

Determine the density of irregular solids as follows:

1. Measure and record the mass of one of the irregular solids.
2. Tie a piece of thread around the solid.
3. Pour some water into a 100 ml graduated cylinder and record the volume.
4. Gently lower the solid into the water, keeping hold of the thread. Record the combined volume of the solid and the water.
5. Dtermine the volume of the solid by subtracting the combined volume from the original volume of the water only.
6. Repeat these steps for the second object.
7. Complete the table below.

Solid	Mass (g)	Volume (ml)	Density (g/ml)
Solid 1			
Solid 2			
Solid 3			

2.6 Summary

- The smallest unit of matter is the **atom**. Atoms can combine to form **molecules**.
- A **molecule** is a group of two or more atoms that are attracted to each other by chemical bonds.
- A **small molecule** consists of a few atoms per molecule. A **giant molecule** consists of millions of atoms per molecule, for example metals and diamonds.
- The structure of a molecule can be represented in a number of ways.
- The **chemical formula** of a molecule is an abbreviated way of showing a molecule, using the symbols for the elements in the molecule. There are two types of chemical formulae: molecular and empirical formula.
- The **molecular formula** of a molecule gives the exact number of atoms of each element that are in the molecule.
- The **empirical formula** of a molecule gives the relative number of atoms of each element in the molecule.
- Molecules can also be represented using **diagrams**.
- A **ball and stick** diagram is a 3-dimensional molecular model that uses 'balls' to represent atoms and 'sticks' to represent the bonds between them.
- A **space-filling model** is also a 3-dimensional molecular model. The atoms are represented by multi-coloured spheres.
- In a molecule, atoms are held together by **chemical bonds** or **intramolecular forces**. Covalent bonds, ionic bonds and metallic bonds are examples of chemical bonds.
- A **covalent bond** exists between non-metal atoms. An **ionic bond** exists between non-metal and metal atoms, and a **metallic bond** exists between metal atoms.
- **Intermolecular forces** are the bonds that hold *molecules* together.
- The **kinetic theory of matter** attempts to explain the behaviour of matter in different phases.
- The theory says that all matter is composed of **particles** which have a certain amount of **energy** which allows them to **move** at different speeds depending on the temperature (energy). There are **spaces** between the particles, and also **attractive forces** between particles when they come close together.

- Understanding chemical bonds, intermolecular forces and the kinetic theory of matter, can help to explain many of the **macroscopic properties** of matter.
- **Melting point** is the temperature at which a *solid* changes its phase to become a *liquid*. The reverse process (change in phase from liquid to solid) is called **freezing**. The stronger the chemical bonds and intermolecular forces in a substance, the higher the melting point will be.
- **Boiling point** is the temperature at which a liquid changes phase to become a gas. The stronger the chemical bonds and intermolecular forces in a substance, the higher the boiling point will be.
- **Density** is a measure of the mass of a substance per unit volume.
- **Viscosity** is a measure of how resistant a liquid is to changing its form.

?

Exercise: Summary exercise

1. Give one word or term for each of the following descriptions.
   (a) The property that determines how easily a liquid flows.
   (b) The change in phase from liquid to gas.
   (c) A composition of two or more atoms that act as a unit.
   (d) Chemical formula that gives the relative number of atoms of each element that are in a molecule.
2. For each of the following questions, choose the one correct answer from the list provided.
   A Ammonia, an ingredient in household cleaners, can be broken down to form one part nitrogen (N) and three parts hydrogen (H). This means that ammonia...
      i. is a colourless gas
      ii. is not a compound
      iii. cannot be an element
      iv. has the formula N_3H
   B If one substance A has a melting point that is *lower* than the melting point of substance B, this suggests that...
      i. A will be a liquid at room temperature.
      ii. The chemical bonds in substance A are weaker than those in substance B.
      iii. The chemical bonds in substance A are stronger than those in substance B.
      iv. B will be a gas at room temperature.
3. Boiling point is an important concept to understand.
   a Define 'boiling point'.
   b What change in phase takes place when a liquid reaches its boiling point?
   c What is the boiling point of water?
   d Use the kinetic theory of matter and your knowledge of intermolecular forces, to explain why water changes phase at this temperature.
4. Refer to the table below which gives the melting and boiling points of a number of elements, and then answer the questions that follow. (*Data from http://www.chemicalelements.com*)

Element	Melting point	Boiling point (^{0}C)
copper	1083	2567
magnesium	650	1107
oxygen	-218.4	-183
carbon	3500	4827
helium	-272	-268.6
sulfur	112.8	444.6

a What state of matter (i.e. solid, liquid or gas) will each of these elements be in at room temperature?

b Which of these elements has the strongest forces between its atoms? Give a reason for your answer.

c Which of these elements has the weakest forces between its atoms? Give a reason for your answer.

Chapter 3

The Atom - Grade 10

We have now looked at many examples of the types of matter and materials that exist around us, and we have investigated some of the ways that materials are classified. But what is it that makes up these materials? And what makes one material different from another? In order to understand this, we need to take a closer look at the building block of matter, the **atom**. Atoms are the basis of all the structures and organisms in the universe. The planets, the sun, grass and trees, the air we breathe, and people are all made up of different combinations of atoms.

3.1 Models of the Atom

It is important to realise that a lot of what we know about the structure of atoms has been developed over a long period of time. This is often how scientific knowledge develops, with one person building on the ideas of someone else. We are going to look at how our modern understanding of the atom has evolved over time.

The idea of atoms was invented by two Greek philosophers, Democritus and Leucippus in the fifth century BC. The Greek word $\alpha\tau o\mu o\nu$ (atom) means *indivisible* because they believed that atoms could not be broken into smaller pieces.

Nowadays, we know that atoms are made up of a *positively charged* **nucleus** in the centre surrounded by *negatively charged* **electrons**. However, in the past, before the structure of the atom was properly understood, scientists came up with lots of different *models* or *pictures* to describe what atoms look like.

Definition: Model
A model is a representation of a system in the real world. Models help us to understand systems and their properties. For example, an *atomic model* represents what the structure of an atom *could* look like, based on what we know about how atoms behave. It is not necessarily a true picture of the exact structure of an atom.

3.1.1 The Plum Pudding Model

After the electron was discovered by J.J. Thomson in 1897, people realised that atoms were made up of even smaller particles than they had previously thought. However, the atomic nucleus had not been discovered yet, and so the 'plum pudding model' was put forward in 1904. In this model, the atom is made up of negative electrons that float in a soup of positive charge, much like plums in a pudding or raisins in a fruit cake (figure 3.1). In 1906, Thomson was awarded the Nobel Prize for his work in this field. However, even with the Plum Pudding Model, there was still no understanding of how these electrons in the atom were arranged.

Figure 3.1: A schematic diagram to show what the atom looked like according to the Plum Pudding model

The discovery of **radiation** was the next step along the path to building an accurate picture of atomic structure. In the early twentieth century, Marie Curie and her husband discovered that some elements (the *radioactive* elements) emit particles, which are able to pass through matter in a similar way to X-rays (read more about this in chapter 7). It was Ernest Rutherford who, in 1911, used this discovery to revise the model of the atom.

3.1.2 Rutherford's model of the atom

Radioactive elements emit different types of particles. Some of these are positively charged alpha (α) particles. Rutherford carried out a series of experiments where he bombarded sheets of gold foil with these particles, to try to get a better understanding of where the positive charge in the atom was. A simplified diagram of his experiment is shown in figure 3.2.

Figure 3.2: Rutherford's gold foil experiment. Figure (a) shows the path of the α particles after they hit the gold sheet. Figure (b) shows the arrangement of atoms in the gold sheets, and the path of the α particles in relation to this.

Rutherford set up his experiment so that a beam of alpha particles was directed at the gold sheets. Behind the gold sheets, was a screen made of zinc sulfide. This screen allowed Rutherford to see where the alpha particles were landing. Rutherford knew that the *electrons* in the gold atoms would not really affect the path of the alpha particles, because the mass of an electron is so much smaller than that of a proton. He reasoned that the positively charged *protons* would be the ones to *repel* the positively charged alpha particles and alter their path.

What he discovered was that most of the alpha particles passed through the foil undisturbed, and could be detected on the screen directly behind the foil (A). Some of the particles ended up being slightly deflected onto other parts of the screen (B). But what was even more interesting was that some of the particles were deflected straight back in the direction from where they had come (C)! These were the particles that had been repelled by the positive protons in the gold atoms. If the Plum Pudding model of the atom were true, then Rutherford would have expected much more repulsion since the positive charge, according to that model, is distributed throughout the atom. But this was not the case. The fact that most particles passed straight through suggested that the positive charge was concentrated in one part of the atom only.

Rutherford's work led to a change in ideas around the atom. His new model described the atom as a tiny, dense, positively charged core called a nucleus, surrounded by lighter, negatively charged electrons. Another way of thinking about this model was that the atom was seen to be like a mini solar system where the electrons orbit the nucleus like planets orbiting around the sun. A simplified picture of this is shown in figure 3.3.

Figure 3.3: Rutherford's model of the atom

3.1.3 The Bohr Model

There were, however, some problems with this model: for example it could not explain the very interesting observation that atoms only emit light at certain wavelengths or frequencies. Niels Bohr solved this problem by proposing that the electrons could only orbit the nucleus in certain special orbits at different energy levels around the nucleus. The exact energies of the orbitals in each energy level depends on the type of atom. Helium for example, has different energy levels to Carbon. If an electron jumps down from a higher energy level to a lower energy level, then light is emitted from the atom. The energy of the light emitted is the same as the gap in the energy between the two energy levels. You can read more about this in section 3.6. The distance between the nucleus and the electron in the lowest energy level of a hydrogen atom is known as the **Bohr radius**.

Interesting Fact

Light has the properties of both a particle *and* a wave! Einstein discovered that light comes in energy packets which are called **photons**. When an electron in an atom changes energy levels, a photon of light is emitted. This photon has the same energy as the difference between the two electron energy levels.

3.2 How big is an atom?

It is difficult sometimes to imagine the size of an atom, or its mass, because we cannot see them, and also because we are not used to working with such small measurements.

3.2.1 How heavy is an atom?

It is possible to determine the mass of a single atom in kilograms. But to do this, you would need very modern *mass spectrometers*, and the values you would get would be very clumsy and difficult to use. The mass of a carbon atom, for example, is about 1.99 x 10^{-26}kg, while the mass of an atom of hydrogen is about 1.67 x 10^{-27}kg. Looking at these very small numbers makes it difficult to compare how much bigger the mass of one atom is when compared to another.

To make the situation simpler, scientists use a different unit of mass when they are describing the mass of an atom. This unit is called the **atomic mass unit** (amu). We can abbreviate (shorten) this unit to just 'u'. If we give carbon an atomic mass of 12 u, then the mass of an atom of hydrogen will be 1 u. You can check this by dividing the mass of a carbon atom in kilograms (see above) by the mass of a hydrogen atom in kilograms (you will need to use a calculator for this!). If you do this calculation, you will see that the mass of a carbon atom is twelve times greater than the mass of a hydrogen atom. When we use atomic mass units instead of kilograms, it becomes easier to see this. Atomic mass units are therefore not giving us the *actual* mass of an atom, but rather its mass *relative* to the mass of other atoms in the Periodic Table. The atomic masses of some elements are shown in table 3.1 below.

Table 3.1: The atomic mass of a number of elements

Element	Atomic mass (u)
Nitrogen (N)	14
Bromine (Br)	80
Magnesium (Mg)	24
Potassium (K)	39
Calcium (Ca)	40
Oxygen (O)	16

The actual value of 1 atomic mass unit is 1.67 x 10^{-24}g or 1.67 x 10^{-27}kg. This is a very tiny mass!

3.2.2 How big is an atom?

pm stands for *picometres*. 1 pm = 10^{-12} m

Atomic diameter also varies depending on the element. On average, the diameter of an atom ranges from 100 pm (Helium) to 670 pm (Caesium). Using different units, 100 pm = 1 Angstrom, and 1 Angstrom = 10^{-10} m. That is the same as saying that 1 Angstrom = 0,0000000010 m or that 100 pm = 0,0000000010 m! In other words, the diameter of an atom ranges from 0.0000000010 m to 0.0000000067 m. This is very small indeed.

3.3 Atomic structure

As a result of the models that we discussed in section 3.1, scientists now have a good idea of what an atom looks like. This knowledge is important because it helps us to understand things like why materials have different properties and why some materials bond with others. Let us now take a closer look at the microscopic structure of the atom.

So far, we have discussed that atoms are made up of a positively charged **nucleus** surrounded by one or more negatively charged **electrons**. These electrons orbit the nucleus.

3.3.1 The Electron

The electron is a very light particle. It has a mass of 9.11 x 10^{-31} kg. Scientists believe that the electron can be treated as a *point particle* or *elementary particle* meaning that it can't be broken down into anything smaller. The electron also carries one unit of **negative** electric charge which is the same as 1.6 x 10^{-19} C (Coulombs).

3.3.2 The Nucleus

Unlike the electron, the nucleus *can* be broken up into smaller building blocks called **protons** and **neutrons**. Together, the protons and neutrons are called **nucleons**.

The Proton

Each proton carries one unit of **positive** electric charge. Since we know that atoms are *electrically neutral*, i.e. do not carry any extra charge, then the number of protons in an atom has to be the same as the number of electrons to balance out the positive and negative charge to zero. The total positive charge of a nucleus is equal to the number of protons in the nucleus. The proton is much heavier than the electron (10 000 times heavier!) and has a mass of 1.6726 x 10^{-27} kg. When we talk about the atomic mass of an atom, we are mostly referring to the combined mass of the protons and neutrons, i.e. the nucleons.

The Neutron

The neutron is electrically neutral i.e. it carries no charge at all. Like the proton, it is much heavier than the electron and its mass is 1.6749 x 10^{-27} kg (slightly heavier than the proton).

Interesting Fact

Rutherford predicted (in 1920) that another kind of particle must be present in the nucleus along with the proton. He predicted this because if there were only positively charged protons in the nucleus, then it should break into bits because of the repulsive forces between the like-charged protons! Also, if protons were the only particles in the nucleus, then a helium nucleus (atomic number 2) would have two protons and therefore only twice the mass of hydrogen. However, it is actually *four* times heavier than hydrogen. This suggested that there must be something else inside the nucleus as well as the protons. To make sure that the atom stays electrically neutral, this particle would have to be neutral itself. In 1932 James Chadwick discovered the neutron and measured its mass.

	proton	**neutron**	**electron**
Mass (kg)	1.6726 x 10^{-27}	1.6749 x 10^{-27}	9.11 x 10^{-31}
Units of charge	+1	0	-1
Charge (C)	1.6 x 10^{-19}	0	-1.6 x 10^{-19}

Table 3.2: Summary of the particles inside the atom

Interesting Fact

Unlike the electron which is thought to be a *point particle* and unable to be broken up into smaller pieces, the proton and neutron **can** be divided. Protons and neutrons are built up of smaller particles called *quarks*. The proton and neutron are made up of 3 quarks each.

3.4 Atomic number and atomic mass number

The chemical properties of an element are determined by the charge of its nucleus, i.e. by the *number of protons*. This number is called the **atomic number** and is denoted by the letter **Z**.

Definition: Atomic number (Z)
The number of protons in an atom

The mass of an atom depends on how many nucleons its nucleus contains. The number of nucleons, i.e. the total number of protons *plus* neutrons, is called the **atomic mass number** and is denoted by the letter **A**.

Definition: Atomic mass number (A)
The number of protons and neutrons in the nucleus of an atom

Standard notation shows the chemical symbol, the atomic mass number and the atomic number of an element as follows:

For example, the iron nucleus which has 26 protons and 30 neutrons, is denoted as

$$^{56}_{26}\mathrm{Fe}\ ,$$

where the total nuclear charge is $Z = 26$ and the mass number $A = 56$. The number of neutrons is simply the difference $N = A - Z$.

Important:
Don't confuse the notation we have used above, with the way this information appears on the Periodic Table. On the Periodic Table, the atomic number usually appears in the top lefthand corner of the block or immediately above the element's symbol. The number below the element's symbol is its **relative atomic mass**. This is not exactly the same as the atomic mass number. This will be explained in section 3.5. The example of iron is used again below.

26
Fe
55.85

You will notice in the example of iron that the atomic mass number is more or less the same as its atomic mass. Generally, an atom that contains n protons and neutrons (i.e. $Z = n$), will have a mass approximately equal to n u. The reason is that a C-12 atom has 6 protons, 6 neutrons and 6 electrons, with the protons and neutrons having about the same mass and the electron mass being negligible in comparison.

Exercise: The structure of the atom

1. Explain the meaning of each of the following terms:
   (a) nucleus
   (b) electron
   (c) atomic mass
2. Complete the following table: (Note: You will see that the atomic masses on the Periodic Table are not *whole numbers*. This will be explained later. For now, you can round off to the nearest whole number.)

Element	**Atomic mass**	**Atomic number**	**Number of protons**	**Number of electrons**	**Number of neutrons**
Mg	24	12			
O			8		
		17			
Ni				28	
	40				20
Zn					
					0
C	12			6	

3. Use standard notation to represent the following elements:
   (a) potassium
   (b) copper
   (c) chlorine
4. For the element $^{35}_{17}$Cl, give the number of ...
   (a) protons
   (b) neutrons
   (c) electrons
   ... in the atom.

5. Which of the following atoms has 7 electrons?
   (a) $^{5}_{2}$He
   (b) $^{13}_{6}$C
   (c) $^{7}_{3}$Li
   (d) $^{15}_{7}$N
6. In each of the following cases, give the number or the element symbol represented by 'X'.
   (a) $^{40}_{18}$X
   (b) $^{x}_{20}$Ca
   (c) $^{31}_{x}$P
7. Complete the following table:

	A	Z	N
$^{235}_{92}$U			
$^{238}_{92}$U			

   In these two different forms of Uranium...
   (a) What is the *same*?
   (b) What is *different*?
   Uranium can occur in different forms, called *isotopes*. You will learn more about isotopes in section 3.5.

3.5 Isotopes

3.5.1 What is an isotope?

If a few neutrons are added to or removed from a nucleus, the chemical properties of the atom will stay the same because its charge is still the same. Therefore, the chemical properties of an element depend on the number of protons inside the atom. This means that such an atom should remain in the same place in the Periodic table. For example, no matter how many neutrons we add or subtract from a nucleus with 6 protons, that element will *always* be called carbon and have the element symbol C (see the Table of Elements). Atoms which have the same number of protons, but a different number of neutrons, are called **isotopes**.

Definition: Isotope
The **isotope** of a particular element, is made up of atoms which have the same number of protons as the atoms in the orginal element, but a different number of neutrons.

The different isotopes of an element have the same atomic number Z but different mass numbers A because they have a different number of neutrons N. The chemical properties of the different isotopes of an element are the same, but they might vary in how stable their nucleus is. Note that if an element is written for example as C-12, the '12' is the atomic mass of that atom. So, Cl-35 has an atomic mass of 35 u, while Cl-37 has an atomic mass of 37 u.

Interesting Fact

In Greek, "same place" reads as *ìσος τòπος* (isos topos). This is why atoms which have the same number of protons, but different numbers of neutrons, are called *isotopes*. They are in the same place on the Periodic Table!

The following worked examples will help you to understand the concept of an isotope better.

Worked Example 1: Isotopes

Question: For the element $^{234}_{92}\mathrm{U}$ (uranium), use standard notation to describe:

1. the isotope with 2 fewer neutrons
2. the isotope with 4 more neutrons

Answer

Step 1 : Go over the definition of isotope

We know that isotopes of any element have the *same* number of protons (same atomic number) in each atom which means that they have the same chemical symbol. However, they have a different number of neutrons, and therefore a different mass number.

Step 2 : Rewrite the notation for the isotopes

Therefore, any isotope of uranium will have the symbol:

$$\mathrm{U}$$

Also, since the number of protons in uranium isotopes is always the same, we can write down the atomic number:

$$_{92}\mathrm{U}$$

Now, if the isotope we want has 2 fewer neutrons than $^{234}_{92}\mathrm{U}$, then we take the original mass number and subtract 2, which gives:

$$^{232}_{92}\mathrm{U}$$

Following the steps above, we can write the isotope with 4 more neutrons as:

$$^{238}_{92}\mathrm{U}$$

Worked Example 2: Isotopes

Question: Which of the following are isotopes of $^{40}_{20}\mathrm{Ca}$?

- $^{40}_{19}\mathrm{K}$
- $^{42}_{20}\mathrm{Ca}$
- $^{40}_{18}\mathrm{Ar}$

Answer

Step 1 : Go over the definition of isotope:

We know that isotopes have the same atomic number but different mass numbers.

Step 2 : Determine which of the elements listed fits the definition of an isotope.

You need to look for the element that has the same atomic number but a different atomic mass number. The only element is $^{12}_{20}$Ca. What is different is that there are 2 more neutrons than in the original element.

Worked Example 3: Isotopes

Question: For the sulfur isotope $^{33}_{16}$S, give the number of...

1. protons
2. nucleons
3. electrons
4. neutrons

Answer

Step 1 : Determine the number of protons by looking at the atomic number, Z.

Z = 16, therefore the number of protons is 16 (answer to (a)).

Step 2 : Determine the number of nucleons by looking at the atomic mass number, A.

A = 33, therefore the number of nucleons is 33 (answer to (b)).

Step 3 : Determine the number of electrons

The atom is neutral, and therefore the number of electrons is the same as the number of protons. The number of electrons is 16 (answer to (c)).

Step 4 : Calculate the number of neutrons

$$N = A - Z = 33 - 16 = 17$$

The number of neutrons is 17 (answer to (d)).

Exercise: Isotopes

1. Atom A has 5 protons and 5 neutrons, and atom B has 6 protons and 5 neutrons. These atoms are...
   (a) allotropes
   (b) isotopes
   (c) isomers
   (d) atoms of different elements
2. For the sulfur isotopes, $^{32}_{16}$S and $^{34}_{16}$S, give the number of...
   (a) protons
   (b) nucleons
   (c) electrons
   (d) neutrons
3. Which of the following are isotopes of Cl^{35}?
   (a) $^{17}_{35}$Cl
   (b) $^{35}_{17}$Cl
   (c) $^{37}_{17}$Cl
4. Which of the following are isotopes of U-235? (X represents an element symbol)
   (a) $^{238}_{92}$X
   (b) $^{238}_{90}$X
   (c) $^{235}_{92}$X

3.5.2 Relative atomic mass

It is important to realise that the atomic mass of isotopes of the same element will be different because they have a different number of nucleons. Chlorine, for example, has two common isotopes which are chlorine-35 and chlorine-37. Chlorine-35 has an atomic mass of 35 u, while chlorine-37 has an atomic mass of 37 u. In the world around us, both of these isotopes occur naturally. It doesn't make sense to say that the element chlorine has an atomic mass of 35 u, or that it has an atomic mass of 37 u. Neither of these are absolutely true since the mass varies depending on the form in which the element occurs. We need to look at how much more common one is than the other in order to calculate the **relative atomic mass** for the element chlorine. This is then the number that will appear on the Periodic Table.

Definition: Relative atomic mass
Relative atomic mass is the average mass of one atom of all the naturally occurring isotopes of a particular chemical element, expressed in atomic mass units.

Worked Example 4: The relative atomic mass of an isotopic element

Question: The element chlorine has two isotopes, chlorine-35 and chlorine-37. The abundance of these isotopes when they occur naturally is 75% chlorine-35 and 25% chlorine-37. Calculate the *average* relative atomic mass for chlorine.

Answer
Step 1 : Calculate the mass contribution of chlorine-35 to the average relative atomic mass

Contribution of Cl-35 = (75/100 x 35) = 26.25 u

Step 2 : Calculate the contribution of chlorine-37 to the average relative atomic mass

Contribution of Cl-37 = (25/100 x 37) = 9.25 u

Step 3 : Add the two values to arrive at the average relative atomic mass of chlorine

Relative atomic mass of chlorine = 26.25 u + 9.25 u = 35.5 u.

If you look on the periodic table, the average relative atomic mass for chlorine is 35,5 u. You will notice that for many elements, the relative atomic mass that is shown is not a whole number. You should now understand that this number is the *average* relative atomic mass for those elements that have naturally occurring isotopes.

Exercise: Isotopes
You are given a sample that contains carbon-12 and carbon-14.

1. Complete the table below:

Isotope	Z	A	Protons	Neutrons	Electrons
Carbon-12					
Carbon-14					
Chlorine-35					
Chlorine-37					

2. If the sample you have contains 90% carbon-12 and 10% carbon-14, calculate the relative atomic mass of an atom in that sample.
3. In another sample, you have 22.5% Cl-37 and 77.5% Cl-35. Calculate the relative atomic mass of an atom in that sample.

Activity :: Group Discussion : The changing nature of scientific knowledge

Scientific knowledge is not static: it changes and evolves over time as scientists build on the ideas of others to come up with revised (and often improved) theories and ideas. In this chapter for example, we saw how peoples' understanding of atomic structure changed as more information was gathered about the atom. There are many more examples like this one in the field of science. Think for example, about our knowledge of the solar system and the origin of the universe, or about the particle and wave nature of light.

Often, these changes in scientific thinking can be very controversial because they disturb what people have come to know and accept. It is important that we realise that what we know *now* about science may also change. An important part of being a scientist is to be a *critical thinker*. This means that you need to question information that you are given and decide whether it is accurate and whether it can be accepted as true. At the same time, you need to learn to be open to new ideas and not to become stuck in what you believe is right... there might just be something new waiting around the corner that you have not thought about!

In groups of 4-5, discuss the following questions:

- Think about some other examples where scientific knowledge has changed because of new ideas and discoveries:
  - What were these new ideas?
  - Were they controversial? If so, why?
  - What role (if any) did *technology* play in developing these new ideas?
  - How have these ideas affected the way we understand the world?
- Many people come up with their own ideas about how the world works. The same is true in science. So how do we, and other scientists, know what to believe and what not to? How do we know when new ideas are 'good' science or 'bad' science? In your groups, discuss some of the things that would need to be done to check whether a new idea or theory was worth listening to, or whether it was not.
- Present your ideas to the rest of the class.

3.6 Energy quantisation and electron configuration

3.6.1 The energy of electrons

You will remember from our earlier discussions, that an atom is made up of a central nucleus, which contains protons and neutrons, and that this nucleus is surrounded by electrons. Although

these electrons all have the same charge and the same mass, each electron in an atom has a different amount of *energy*. Electrons that have the *lowest* energy are found closest to the nucleus where the attractive force of the positively charged nucleus is the greatest. Those electrons that have *higher* energy, and which are able to overcome the attractive force of the nucleus, are found further away.

3.6.2 Energy quantisation and line emission spectra

If the energy of an atom is increased (for example when a substance is heated), the energy of the electrons inside the atom can be increased (when an electron has a higher energy than normal it is said to be "excited"). For the excited electron to go back to its original energy (called the ground state), it needs to release energy. It releases energy by emitting light. If one heats up different elements, one will see that for each element, light is emitted only at certain frequencies (or wavelengths). Instead of a smooth continuum of frequencies, we see lines (called emission lines) at particular frequencies. These frequencies correspond to the energy of the emitted light. If electrons could be excited to any energy and lose any amount of energy, there would be a continuous spread of light frequencies emitted. However, the sharp lines we see mean that there are only certain particular energies that an electron can be excited to, or can lose, for each element.

You can think of this like going up a flight of steps: you can't lift your foot by *any* amount to go from the ground to the first step. If you lift your foot too low you'll bump into the step and be stuck on the ground level. You have to lift your foot just the right amount (the height of the step) to go to the next step, and so on. The same goes for electrons and the amount of energy they can have. This is called **quantisation of energy** because there are only certain quantities of energy that an electron can have in an atom. Like steps, we can think of these quantities as **energy levels** in the atom. The energy of the light released when an electron drops down from a higher energy level to a lower energy level is the same as the difference in energy between the two levels.

3.6.3 Electron configuration

Electrons are arranged in energy levels around the nucleus of an atom. Electrons that are in the energy level that is closest to the nucleus, will have the lowest energy and those further away will have a higher energy. Each energy level can only hold a certain number of electrons, and an electron will only be found in the second energy level once the first energy level is full. The same rule applies for the higher energy levels. You will need to learn the following rules:

- The 1st energy level can hold a maximum of 2 electrons
- The 2nd energy level can hold a maximum of 8 electrons
- The 3rd energy level can hold a maximum of 8 electrons
- If the number of electrons in the atom is greater than 18, they will need to move to the 4th energy level.

In the following examples, the energy levels are shown as concentric circles around the central nucleus.

1. **Lithium**

   Lithium (Li) has an atomic number of 3, meaning that in a neutral atom, the number of electrons will also be 3. The first two electrons are found in the first energy level, while the third electron is found in the second energy level (figure 3.11).

2. **Fluorine**

   Fluorine (F) has an atomic number of 9, meaning that a neutral atom also has 9 electrons. The first 2 electrons are found in the first energy level, while the other 7 are found in the second energy level (figure 3.12).

Figure 3.4: The arrangement of electrons in a lithium atom.

Figure 3.5: The arrangement of electrons in a fluorine atom.

3. **Argon**

   Argon has an atomic number of 18, meaning that a neutral atom also has 18 electrons. The first 2 electrons are found in the first energy level, the next 8 are found in the second energy level, and the last 8 are found in the third energy level (figure 3.6).

Figure 3.6: The arrangement of electrons in an argon atom.

But the situation is slightly more complicated than this. Within each energy level, the electrons move in **orbitals**. An orbital defines the spaces or regions where electrons move.

Definition: Atomic orbital
An atomic orbital is the region in which an electron may be found around a single atom.

There are different orbital shapes, but we will be dealing with only two. These are the 's' and 'p' orbitals (there are also 'd' and 'f' orbitals). The first energy level contains only one 's' orbital, the second energy level contains one 's' orbital and three 'p' orbitals and the third energy level also contains one 's' orbital and three 'p' orbitals. Within each energy level, the 's' orbital is at a lower energy than the 'p' orbitals. This arrangement is shown in figure 3.7.

When we want to show how electrons are arranged in an atom, we need to remember the following principles:

Figure 3.7: The positions of the first ten orbits of an atom on an energy diagram. Note that each block is able to hold two electrons.

- Each orbital can only hold **two electrons**. Electrons that occur together in an orbital are called an **electron pair**. These electrons spin in opposite directions around their own axes.
- An electron will always try to enter an orbital with the lowest possible energy.
- An electron will occupy an orbital on its own, rather than share an orbital with another electron. An electron would also rather occupy a lower energy orbital *with* another electron, before occupying a higher energy orbital. In other words, within one energy level, electrons will fill an 's' orbital before starting to fill 'p' orbitals.

The way that electrons are arranged in an atom is called its **electron configuration**.

Definition: Electron configuration
Electron configuration is the arrangement of electrons in an atom, molecule, or other physical structure.

An element's electron configuration can be represented using **Aufbau diagrams** or energy level diagrams. An Aufbau diagram uses arrows to represent electrons. You can use the following steps to help you to draw an Aufbau diagram:

1. Determine the number of electrons that the atom has.
2. Fill the 's' orbital in the first energy level (the 1s orbital) with the first two electrons.
3. Fill the 's' orbital in the second energy level (the 2s orbital) with the second two electrons.
4. Put one electron in each of the three 'p' orbitals in the second energy level (the 2p orbitals), and then if there are still electrons remaining, go back and place a second electron in each of the 2p orbitals to complete the electron pairs.
5. Carry on in this way through each of the successive energy levels until all the electrons have been drawn.

Important:
When there are two electrons in an orbital, the electrons are called an **electron pair**. If the orbital only has one electron, this electron is said to be an **unpaired electron**. Electron pairs are shown with arrows in opposite directions. This is because when two electrons occupy the same orbital, they spin in opposite directions on their axes.

An Aufbau diagram for the element Lithium is shown in figure 3.8.

Figure 3.8: The electron configuration of Lithium, shown on an Aufbau diagram

A special type of notation is used to show an atom's electron configuration. The notation describes the energy levels, orbitals and the number of electrons in each. For example, the electron configuration of lithium is $1s^2\ 2s^1$. The number and letter describe the energy level and orbital, and the number above the orbital shows how many electrons are in that orbital.

Aufbau diagrams for the elements fluorine and argon are shown in figures 3.9 and 3.10 respectively. Using standard notation, the electron configuration of fluorine is $1s^2\ 2s^2\ 2p^5$ and the electron configuration of argon is $1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^6$.

Figure 3.9: An Aufbau diagram showing the electron configuration of fluorine

3.6.4 Core and valence electrons

Electrons in the outermost energy level of an atom are called **valence electrons**. The electrons that are in the energy shells closer to the nucleus are called **core electrons**. Core electrons are all the electrons in an atom, excluding the valence electrons. An element that has its valence energy level full is *more stable* and *less likely to react* than other elements with a valence energy level that is not full.

Definition: Valence electrons
The electrons in the outer energy level of an atom

Definition: Core electrons
All the electrons in an atom, excluding the valence electrons

3.6.5 The importance of understanding electron configuration

By this stage, you may well be wondering why it is important for you to understand how electrons are arranged around the nucleus of an atom. Remember that during chemical reactions, when atoms come into contact with one another, it is the *electrons* of these atoms that will interact first. More specifically, it is the **valence electrons** of the atoms that will determine how they

Argon

Figure 3.10: An Aufbau diagram showing the electron configuration of argon

react with one another.

To take this a step further, an atom is at its most stable (and therefore *unreactive*) when all its orbitals are full. On the other hand, an atom is least stable (and therefore most *reactive*) when its valence electron orbitals are not full. This will make more sense when we go on to look at chemical bonding in a later chapter. To put it simply, the valence electrons are largely responsible for an element's chemical behaviour, and elements that have the same number of valence electrons often have similar chemical properties.

Exercise: Energy diagrams and electrons

1. Draw Aufbau diagrams to show the electron configuration of each of the following elements:
   (a) magnesium
   (b) potassium
   (c) sulfur
   (d) neon
   (e) nitrogen
2. Use the Aufbau diagrams you drew to help you complete the following table:

Element	No. of energy levels	No. of core electrons	No. of valence electrons	Electron configuration (standard notation)
Mg				
K				
S				
Ne				
N				

3. Rank the elements used above in order of *increasing reactivity*. Give reasons for the order you give.

Activity :: Group work : Building a model of an atom

Earlier in this chapter, we talked about different 'models' of the atom. In science, one of the uses of models is that they can help us to understand the structure of something that we can't see. In the case of the atom, models help us to build a picture in our heads of what the atom looks like.

Models are often simplified. The small toy cars that you may have played with as a child are models. They give you a good idea of what a real car looks like, but they are much smaller and much simpler. A model cannot always be absolutely accurate and it is important that we realise this so that we don't build up a false idea about something.

In groups of 4-5, you are going to build a model of an atom. Before you start, think about these questions:

- What information do I know about the structure of the atom? (e.g. what parts make it up? how big is it?)
- What materials can I use to represent these parts of the atom as accurately as I can?
- How will I put all these different parts together in my model?

As a group, share your ideas and then plan how you will build your model. Once you have built your model, discuss the following questions:

- Does our model give a good idea of what the atom actually looks like?
- In what ways is our model *inaccurate*? For example, we know that electrons *move* around the atom's nucleus, but in your model, it might not have been possible for you to show this.
- Are there any ways in which our model could be improved?

Now look at what other groups have done. Discuss the same questions for each of the models you see and record your answers.

3.7 Ionisation Energy and the Periodic Table

3.7.1 Ions

In the previous section, we focused our attention on the electron configuration of *neutral* atoms. In a neutral atom, the number of protons is the same as the number of electrons. But what

happens if an atom *gains* or *loses* electrons? Does it mean that the atom will still be part of the same element?

A change in the number of electrons of an atom does not change the type of atom that it is. However, the *charge* of the atom will change. If electrons are added, then the atom will become *more negative*. If electrons are taken away, then the atom will become *more positive*. The atom that is formed in either of these cases is called an **ion**. Put simply, an ion is a charged atom.

Definition: Ion
An ion is a charged atom. A positively charged ion is called a **cation** e.g. Na^+, and a negatively charged ion is called an **anion** e.g. F^-. The charge on an ion depends on the number of electrons that have been lost or gained.

Look at the following examples. Notice the number of valence electrons in the neutral atom, the number of electrons that are lost or gained, and the final charge of the ion that is formed.

Lithium

A lithium atoms loses one electrons to form a positive ion (figure 3.11).

Figure 3.11: The arrangement of electrons in a lithium ion.

In this example, the lithium atom loses an electron to form the cation Li^+.

Fluorine

A fluorine atom gains one electron to form a negative ion (figure 3.12).

Figure 3.12: The arrangement of electrons in a fluorine ion.

Activity :: Investigation : The formation of ions

1. Use the diagram for lithium as a guide and draw similar diagrams to show how each of the following ions is formed:
   (a) Mg^{2+}
   (b) Na^{+}
   (c) Cl^{-}
   (d) O^{2-}
2. Do you notice anything interesting about the charge on each of these ions? Hint: Look at the number of valence electrons in the neutral atom and the charge on the final ion.

Observations:

Once you have completed the activity, you should notice that:

- In each case the number of electrons that is either gained or lost, is the same as the number of electrons that are needed for the atoms to achieve a full or an empty valence energy level.
- If you look at an energy level diagram for sodium (Na), you will see that in a neutral atom, there is only one valence electron. In order to achieve an empty valence level, and therefore a more stable state for the atom, this electron will be *lost*.
- In the case of oxygen (O), there are six valence electrons. To fill the valence energy level, it makes more sense for this atom to *gain* two electrons. A negative ion is formed.

3.7.2 Ionisation Energy

Ionisation energy is the energy that is needed to remove one electron from an atom. The ionisation energy will be different for different atoms.

The second ionisation energy is the energy that is needed to remove a second electron from an atom, and so on. As an energy level becomes more full, it becomes more and more difficult to remove an electron and the ionisation energy *increases*. On the Periodic Table of the Elements, a *group* is a vertical column of the elements, and a *period* is a horizontal row. In the periodic table, ionisation energy *increases* across a period, but *decreases* as you move down a group. The lower the ionisation energy, the more reactive the element will be because there is a greater chance of electrons being involved in chemical reactions. We will look at this in more detail in the next section.

?

Exercise: The formation of ions

Match the information in column A with the information in column B by writing only the letter (A to I) next to the question number (1 to 7)

1. A positive ion that has 3 less electrons than its neutral atom	A. Mg^{2+}
2. An ion that has 1 more electron than its neutral atom	B. Cl^{-}
3. The anion that is formed when bromine gains an electron	C. CO_3^{2-}
4. The cation that is formed from a magnesium atom	D. Al^{3+}
5. An example of a compound ion	E. Br^{2-}
6. A positive ion with the electron configuration of argon	F. K^{+}
7. A negative ion with the electron configuration of neon	G. Mg^{+}
	H. O^{2-}
	I. Br^{-}

3.8 The Arrangement of Atoms in the Periodic Table

The **periodic table of the elements** is a tabular method of showing the chemical elements. Most of the work that was done to arrive at the periodic table that we know, can be attributed to a man called **Dmitri Mendeleev** in 1869. Mendeleev was a Russian chemist who designed the table in such a way that recurring ("periodic") trends in the properties of the elements could be shown. Using the trends he observed, he even left gaps for those elements that he thought were 'missing'. He even predicted the properties that he thought the missing elements would have when they were discovered. Many of these elements were indeed discovered and Mendeleev's predictions were proved to be correct.

To show the recurring properties that he had observed, Mendeleev began new rows in his table so that elements with similar properties were in the same vertical columns, called **groups**. Each row was referred to as a **period**. One important feature to note in the periodic table is that all the non-metals are to the right of the zig-zag line drawn under the element boron. The rest of the elements are metals, with the exception of hydrogen which occurs in the first block of the table despite being a non-metal.

Figure 3.13: A simplified diagram showing part of the Periodic Table

3.8.1 Groups in the periodic table

A *group* is a vertical column in the periodic table, and is considered to be the most important way of classifying the elements. If you look at a periodic table, you will see the groups numbered

at the top of each column. The groups are numbered from left to right as follows: 1, 2, then an open space which contains the **transition elements**, followed by groups 3 to 8. These numbers are normally represented using roman numerals. In some periodic tables, all the groups are numbered from left to right from number 1 to number 18. In some groups, the elements display very similar chemical properties, and the groups are even given separate names to identify them.

The characteristics of each group are mostly determined by the electron configuration of the atoms of the element.

- *Group 1:* These elements are known as the **alkali metals** and they are very reactive.

Figure 3.14: Electron diagrams for some of the Group 1 elements

---

Activity :: Investigation : The properties of elements
Refer to figure 3.14.

1. Use a Periodic Table to help you to complete the last two diagrams for sodium (Na) and potassium (K).
2. What do you notice about the number of electrons in the valence energy level in each case?
3. Explain why elements from group 1 are more reactive than elements from group 2 on the periodic table (Hint: Think back to 'ionisation energy').

- *Group 2:* These elements are known as the **alkali earth metals**. Each element only has two valence electrons and so in chemical reactions, the group 2 elements tend to *lose* these electrons so that the energy shells are complete. These elements are less reactive than those in group 1 because it is more difficult to lose two electrons than it is to lose one. *Group 3* elements have three valence electrons.

Important: The number of valence electrons of an element corresponds to its group number on the periodic table.

- *Group 7:* These elements are known as the **halogens**. Each element is missing just one electron from its outer energy shell. These elements tend to *gain* electrons to fill this shell, rather than losing them.
- *Group 8:* These elements are the **noble gases**. All of the energy shells of the halogens are full, and so these elements are very unreactive.
- *Transition metals:* The differences between groups in the transition metals are not usually dramatic.

Helium Lithium

Figure 3.15: Electron diagrams for two of the noble gases, helium (He) and neon (Ne).

It is worth noting that in each of the groups described above, the **atomic diameter** of the elements increases as you move down the group. This is because, while the number of valence electrons is the same in each element, the number of core electrons increases as one moves down the group.

3.8.2 Periods in the periodic table

A **period** is a horizontal row in the periodic table of the elements. Some of the trends that can be observed within a period are highlighted below:

- As you move from one group to the next within a period, the number of valence electrons increases by one each time.
- Within a single period, all the valence electrons occur in the same energy shell. If the period increases, so does the energy shell in which the valence electrons occur.
- In general, the diameter of atoms decreases as one moves from left to right across a period. Consider the attractive force between the positively charged nucleus and the negatively charged electrons in an atom. As you move across a period, the number of protons in each atom increases. The number of electrons also increases, but these electrons will still be in the same energy shell. As the number of protons increases, the force of attraction between the nucleus and the electrons will increase and the atomic diameter will decrease.
- Ionisation energy increases as one moves from left to right across a period. As the valence electron shell moves closer to being full, it becomes more difficult to remove electrons. The opposite is true when you move down a *group* in the table because more energy shells are being added. The electrons that are closer to the nucleus 'shield' the outer electrons from the attractive force of the positive nucleus. Because these electrons are not being held to the nucleus as strongly, it is easier for them to be removed and the ionisation energy decreases.
- In general, the reactivity of the elements decreases from left to right across a period.

?

Exercise: Trends in ionisation energy

Refer to the data table below which gives the ionisation energy (in kJ/mol) and atomic number (Z) for a number of elements in the periodic table:

Z	Ionisation energy	Z	Ionisation energy
1	1310	10	2072
2	2360	11	494
3	517	12	734
4	895	13	575
5	797	14	783
6	1087	15	1051
7	1397	16	994
8	1307	17	1250
9	1673	18	1540

1. Draw a line graph to show the relationship between atomic number (on the x-axis) and ionisation energy (y-axis).
2. Describe any trends that you observe.
3. Explain why...
   (a) the ionisation energy for Z=2 is higher than for Z=1
   (b) the ionisation energy for Z=3 is lower than for Z=2
   (c) the ionisation energy increases between Z=5 and Z=7

?

Exercise: Elements in the Periodic Table
Refer to the elements listed below:

Lithium (Li); Chlorine (Cl); Magnesium (Mg); Neon (Ne); Oxygen (O); Calcium (Ca); Carbon (C)

Which of the elements listed above:

1. belongs to Group 1
2. is a halogen
3. is a noble gas
4. is an alkali metal
5. has an atomic number of 12
6. has 4 neutrons in the nucleus of its atoms
7. contains electrons in the 4th energy level
8. has only one valence electron
9. has all its energy orbitals full
10. will have chemical properties that are most similar
11. will form positive ions

3.9 Summary

- Much of what we know today about the atom, has been the result of the work of a number of scientists who have added to each other's work to give us a good understanding of atomic structure.

- Some of the important scientific contributors include **J.J.Thomson** (discovery of the electron, which led to the Plum Pudding Model of the atom), **Ernest Rutherford** (discovery that positive charge is concentrated in the centre of the atom) and **Niels Bohr** (the arrangement of electrons around the nucleus in energy levels).
- Because of the very small mass of atoms, their mass in measured in **atomic mass units** (u). 1 u = 1.67×10^{-24} g.
- An atom is made up of a central **nucleus** (containing **protons** and **neutrons**), surrounded by **electrons**.
- The **atomic number** (Z) is the number of protons in an atom.
- The **atomic mass number** (A) is the number of protons and neutrons in the nucleus of an atom.
- The **standard notation** that is used to write an element, is $^{A}_{Z}$X, where X is the element symbol, A is the atomic mass number and Z is the atomic number.
- The **isotope** of a particular element is made up of atoms which have the same number of protons as the atoms in the original element, but a different number of neutrons. This means that not all atoms of an element will have the same atomic mass.
- The **relative atomic mass** of an element is the average mass of one atom of all the naturally occurring isotopes of a particular chemical element, expressed in atomic mass units. The relative atomic mass is written under the elements' symbol on the Periodic Table.
- The energy of electrons in an atom is **quantised**. Electrons occur in specific energy levels around an atom's nucleus.
- Within each energy level, an electron may move within a particular shape of **orbital**. An orbital defines the space in which an electron is most likely to be found. There are different orbital shapes, including s, p, d and f orbitals.
- Energy diagrams such as **Aufbau diagrams** are used to show the electron configuration of atoms.
- The electrons in the outermost energy level are called **valence electrons**.
- The electrons that are not valence electrons are called **core electrons**.
- Atoms whose outermost energy level is full, are less chemically reactive and therefore more stable, than those atoms whose outer energy level is not full.
- An **ion** is a charged atom. A **cation** is a positively charged ion and an **anion** is a negatively charged ion.
- When forming an ion, an atom will lose or gain the number of electrons that will make its valence energy level full.
- An element's **ionisation energy** is the energy that is needed to remove one electron from an atom.
- Ionisation energy increases across a **period** in the Periodic Table.
- Ionisation energy decreases down a **group** in the Periodic Table.

Exercise: Summary

1. Write down only the word/term for each of the following descriptions.
   (a) The sum of the number of protons and neutrons in an atom

(b) The defined space around an atom's nucleus, where an electron is most likely to be found

2. For each of the following, say whether the statement is True or False. If it is False, re-write the statement correctly.
   (a) $^{20}_{10}$Ne and $^{22}_{10}$Ne each have 10 protons, 12 electrons and 12 neutrons.
   (b) The atomic mass of any atom of a particular element is always the same.
   (c) It is safer to use helium gas rather than hydrogen gas in balloons.
   (d) Group 1 elements readily form negative ions.
3. Multiple choice questions: In each of the following, choose the **one** correct answer.
   (a) The three basic components of an atom are:
      i. protons, neutrons, and ions
      ii. protons, neutrons, and electrons
      iii. protons, neutrinos, and ions
      iv. protium, deuterium, and tritium
   (b) The charge of an atom is...
      i. positive
      ii. neutral
      iii. negative
   (c) If Rutherford had used neutrons instead of alpha particles in his scattering experiment, the neutrons would...
      i. not deflect because they have no charge
      ii. have deflected more often
      iii. have been attracted to the nucleus easily
      iv. have given the same results
   (d) Consider the isotope $^{234}_{92}$U. Which of the following statements is *true*?
      i. The element is an isotope of $^{234}_{94}$Pu
      ii. The element contains 234 neutrons
      iii. The element has the same electron configuration as $^{238}_{92}$U
      iv. The element has an atomic mass number of 92
   (e) The electron configuration of an atom of chlorine can be represented using the following notation:
      i. $1s^2\ 2s^8\ 3s^7$
      ii. $1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^5$
      iii. $1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^6$
      iv. $1s^2\ 2s^2\ 2p^5$
4. The following table shows the first ionisation energies for the elements of period 1 and 2.

Period	Element	First ionisation energy ($kJ.mol^{-1}$)
1	H	1312
	He	2372
2	Li	520
	Be	899
	B	801
	C	1086
	N	1402
	O	1314
	F	1681
	Ne	2081

   (a) What is the meaning of the term *first ionisation energy*?
   (b) Identify the pattern of first ionisation energies in a period.
   (c) Which TWO elements exert the strongest attractive forces on their electrons? Use the data in the table to give a reason for your answer.

(d) Draw Aufbau diagrams for the TWO elements you listed in the previous question and explain why these elements are so stable.

(e) It is safer to use helium gas than hydrogen gas in balloons. Which property of helium makes it a safer option?

(f) 'Group 1 elements readily form positive ions'.
Is this statement correct? Explain your answer by referring to the table.

Chapter 4

Atomic Combinations - Grade 11

When you look at the matter around you, you will realise that atoms seldom exist on their own. More often, the things around us are made up of different atoms that have been joined together. This is called **chemical bonding**. Chemical bonding is one of the most important processes in chemistry because it allows all sorts of different molecules and combinations of atoms to form, which then make up the objects in the complex world around us. There are, however, some atoms that *do* exist on their own, and which do not bond with others. The **noble gases** in Group 8 of the Periodic Table, behave in this way. They include elements like neon (Ne), helium (He) and argon (Ar). The important question then is, why do some atoms bond but others do not?

4.1 Why do atoms bond?

As we begin this section, it's important to remember that what we will go on to discuss is a *model* of bonding, that is based on a particular *model* of the atom. You will remember from section 3.1 that a model is a *representation* of what is happening in reality. In the model of the atom that has been used so far, the atom is made up of a central nucleus, surrounded by electrons that are arranged in fixed energy levels (also sometimes called *shells*). Within each energy level, electrons move in *orbitals* of different shapes. The electrons in the outermost energy level of an atom are called the **valence electrons**. This model of the atom is useful in trying to understand how different types of bonding take place between atoms.

You will remember from these earlier discussions of electrons and energy levels in the atom, that electrons always try to occupy the *lowest* possible energy level. In the same way, an atom also prefers to exist at the lowest possible energy state so that it is most *stable*. An atom is most stable when all its valence electron orbitals are *full*. In other words, the outer energy level of the atom contains the maximum number of electrons that it can. A stable atom is also an *unreactive* one, and is unlikely to bond with other atoms. This explains why the noble gases are unreactive and why they exist as atoms, rather than as molecules. Look for example at the electron configuration of neon ($1s^2\ 2s^2\ 3p^6$). Neon has eight valence electrons in its valence energy shell. This is the maximum that it can hold and so neon is very stable and unreactive, and will not form new bonds. Other atoms, whose valence energy levels are not full, are more likely to bond in order to become more stable. We are going to look a bit more closely at some of the energy changes that take place when atoms bond.

4.2 Energy and bonding

Let's start by imagining that there are two hydrogen atoms approaching one another. As they move closer together, there are three forces that act on the atoms at the same time. These forces are shown in figure 4.1 and are described below:

Figure 4.1: Forces acting on two approaching atoms: (1) repulsion between electrons, (2) attraction between protons and electrons and (3) repulsion between protons.

1. **repulsive force** between the electrons of the atoms, since like charges repel
2. **attractive force** between the nucleus of one atom and the electrons of another
3. **repulsive force** between the two positively-charged nuclei

Now look at figure 4.2 to understand the energy changes that take place when the two atoms move towards each other.

Figure 4.2: Graph showing the change in energy that takes place as atoms move closer together

In the example of the two hydrogen atoms, where the resultant force between them is attraction, the energy of the system is zero when the atoms are far apart (point A), because there is no interaction between the atoms. When the atoms are closer together, attractive forces dominate and the atoms are pulled towards each other. As this happens, the *potential energy* of the system decreases because energy would now need to be *supplied* to the system in order to move the atoms apart. However, as the atoms move closer together (i.e. *left* along the horizontal axis of the graph), repulsive forces start to dominate and this causes the potential energy of the system to rise again. At some point, the attractive and repulsive effects are balanced, and the energy of the system is at its minimum (point X). It is at this point, when the energy is at a minimum, that bonding takes place.

The distance marked 'P' is the **bond length**, i.e. the distance between the nuclei of the atoms when they bond. 'Q' represents the **bond energy** i.e. the amount of energy that must be added to the system to break the bonds that have formed. **Bond strength** means how strongly one atom attracts and is held to another. The strength of a bond is related to the bond length, the size of the bonded atoms and the number of bonds between the atoms. In general, the shorter the bond length, the stronger the bond between the atoms, and the smaller the atoms involved, the stronger the bond. The greater the number of bonds between atoms, the greater will be the bond strength.

4.3 What happens when atoms bond?

A **chemical bond** is formed when atoms are held together by attractive forces. This attraction occurs when electrons are *shared* between atoms, or when electrons are *exchanged* between the atoms that are involved in the bond. The sharing or exchange of electrons takes place so that the outer energy levels of the atoms involved are filled and the atoms are more stable. If an electron is **shared**, it means that it will spend its time moving in the electron orbitals around *both* atoms. If an electron is **exchanged** it means that it is transferred from one atom to another, in other words one atom *gains* an electron while the other *loses* an electron.

Definition: Chemical bond
A chemical bond is the physical process that causes atoms and molecules to be attracted to each other, and held together in more stable chemical compounds.

The type of bond that is formed depends on the elements that are involved. In this section, we will be looking at three types of chemical bonding: **covalent**, **ionic** and **metallic bonding**.

You need to remember that it is the *valence electrons* that are involved in bonding and that atoms will try to fill their outer energy levels so that they are more stable.

4.4 Covalent Bonding

4.4.1 The nature of the covalent bond

Covalent bonding occurs between the atoms of **non-metals**. The outermost orbitals of the atoms overlap so that unpaired electrons in each of the bonding atoms can be shared. By overlapping orbitals, the outer energy shells of all the bonding atoms are filled. The shared electrons move in the orbitals around *both* atoms. As they move, there is an attraction between these negatively charged electrons and the positively charged nuclei, and this force holds the atoms together in a covalent bond.

Definition: Covalent bond
Covalent bonding is a form of chemical bonding where pairs of electrons are shared between atoms.

Below are a few examples. Remember that it is only the *valence electrons* that are involved in bonding, and so when diagrams are drawn to show what is happening during bonding, it is only these electrons that are shown. Circles and crosses represent electrons in different atoms.

Worked Example 5: Covalent bonding

Question: How do hydrogen and chlorine atoms bond covalently in a molecule of hydrogen chloride?

Answer
Step 1 : Determine the electron configuration of each of the bonding atoms.
A chlorine atom has 17 electrons, and an electron configuration of $1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^5$. A hydrogen atom has only 1 electron, and an electron configuration of $1s^1$.

Step 2 : Determine the number of valence electrons for each atom, and how many of the electrons are paired or unpaired.
Chlorine has 7 valence electrons. One of these electrons is unpaired. Hydrogen has 1 valence electron and it is unpaired.

Step 3 : Look to see how the electrons can be shared between the atoms so that the outermost energy levels of both atoms are full.
The hydrogen atom needs one more electron to complete its valence shell. The chlorine atom also needs one more electron to complete its shell. Therefore *one pair of electrons* must be shared between the two atoms. In other words, one electron from the chlorine atom will spend some of its time orbiting the hydrogen atom so that hydrogen's valence shell is full. The hydrogen electron will spend some of its time orbiting the chlorine atom so that chlorine's valence shell is also full. A molecule of hydrogen chloride is formed (figure 4.3). Notice the shared electron pair in the overlapping orbitals.

Figure 4.3: Covalent bonding in a molecule of hydrogen chloride

Worked Example 6: Covalent bonding involving multiple bonds

Question: How do nitrogen and hydrogen atoms bond to form a molecule of ammonia (NH_3)?

Answer
Step 1 : Determine the electron configuration of each of the bonding atoms.
A nitrogen atom has 7 electrons, and an electron configuration of $1s^2\ 2s^2\ 2p^3$. A hydrogen atom has only 1 electron, and an electron configuration of $1s^1$.

Step 2 : Determine the number of valence electrons for each atom, and how many of the electrons are paired or unpaired.
Nitrogen has 5 valence electrons meaning that 3 electrons are unpaired. Hydrogen has 1 valence electron and it is unpaired.

Step 3 : Look to see how the electrons can be shared between the atoms so that the outer energy shells of all atoms are full.

Each hydrogen atom needs one more electron to complete its valence energy shell. The nitrogen atom needs three more electrons to complete its valence energy shell. Therefore *three pairs of electrons* must be shared between the four atoms involved. The nitrogen atom will share three of its electrons so that each of the hydrogen atoms now has a complete valence shell. Each of the hydrogen atoms will share its electron with the nitrogen atom to complete its valence shell (figure 4.4).

Figure 4.4: Covalent bonding in a molecule of ammonia

The above examples all show **single covalent bonds**, where only one pair of electrons is shared between *the same two atoms*. If two pairs of electrons are shared between the same two atoms, this is called a **double bond**. A **triple bond** is formed if three pairs of electrons are shared.

Worked Example 7: Covalent bonding involving a double bond

Question: How do oxygen atoms bond covalently to form an oxygen molecule?

Answer
Step 1 : Determine the electron configuration of the bonding atoms.
Each oxygen atom has 8 electrons, and their electron configuration is $1s^2\ 2s^2\ 2p^4$.

Step 2 : Determine the number of valence electrons for each atom and how many of these electrons are paired and unpaired.
Each oxygen atom has 6 valence electrons, meaning that each atom has 2 unpaired electrons.

Step 3 : Look to see how the electrons can be shared between atoms so that the outer energy shells of all the atoms are full.
Each oxygen atom needs two more electrons to complete its valence energy shell. Therefore *two pairs of electrons* must be shared between the two oxygen atoms so that both valence shells are full. Notice that the two electron pairs are being shared between *the same two* atoms, and so we call this a **double bond** (figure 4.5).

You will have noticed in the above examples that the number of electrons that are involved in bonding varies between atoms. We say that the **valency** of the atoms is different.

Definition: Valency
The number of electrons in an atom which are used to form a bond.

Figure 4.5: A double covalent bond in an oxygen molecule

In the first example, the valency of both hydrogen and chlorine is one, therefore there is a single covalent bond between these two atoms. In the second example, nitrogen has a valency of three and hydrogen has a valency of one. This means that three hydrogen atoms will need to bond with a single nitrogen atom. There are three *single* covalent bonds in a molecule of ammonia. In the third example, the valency of oxygen is two. This means that each oxygen atom will form two bonds with another atom. Since there is only one other atom in a molecule of O_2, a *double covalent* bond is formed between these two atoms.

Important: There is a relationship between the valency of an element and its position on the Periodic Table. For the elements in groups 1 to 4, the valency is the same as the group number. For elements in groups 5 to 7, the valency is calculated by subtracting the group number from 8. For example, the valency of fluorine (group 7) is 8-7=1, while the valency of calcium (group 2) is 2. Some elements have more than one possible valency, so you always need to be careful when you are writing a chemical formula. Often, the valency will be written in a bracket after the element symbol e.g. carbon (iv) oxide, means that in this molecule carbon has a valency of 4.

Exercise: Covalent bonding and valency

1. Explain the difference between the *valence electrons* and the *valency* of an element.
2. Complete the table below by filling in the number of valence electrons and the valency for each of the elements shown:

Element	**No. of valence electrons**	**No. of electrons needed to fill outer shell**	**Valency**
Mg			
K			
F			
Ar			
C			
N			
O			

3. Draw simple diagrams to show how electrons are arranged in the following covalent molecules:
   (a) Calcium oxide (CaO)
   (b) Water (H_2O)
   (c) Chlorine (Cl_2)

4.5 Lewis notation and molecular structure

Although we have used diagrams to show the structure of molecules, there are other forms of notation that can be used, such as **Lewis notation** and **Couper notation**. **Lewis notation** uses dots and crosses to represent the **valence electrons** on different atoms. The chemical symbol of the element is used to represent the nucleus and the core electrons of the atom.

So, for example, a hydrogen atom would be represented like this:

A chlorine atom would look like this:

A molecule of hydrogen chloride would be shown like this:

The dot and cross inbetween the two atoms, represent the pair of electrons that are shared in the covalent bond.

Worked Example 8: Lewis notation: Simple molecules

Question: Represent the molecule H_2O using Lewis notation

Answer

Step 1 : For each atom, determine the number of valence electrons in the atom, and represent these using dots and crosses.

The electron configuration of hydrogen is $1s^1$ and the electron configuration for oxygen is $1s^2\ 2s^2\ 2p^4$. Each hydrogen atom has one valence electron, which is unpaired, and the oxygen atom has six valence electrons with two unpaired.

Step 2 : Arrange the electrons so that the outermost energy level of each atom is full.

The water molecule is represented below.

Worked Example 9: Lewis notation: Molecules with multiple bonds

Question: Represent the molecule HCN using Lewis notation

Answer

Step 1 : For each atom, determine the number of valence electrons that the atom has from its electron configuration.

The electron configuration of hydrogen is $1s^1$, the electron configuration of nitrogen is $1s^2\ 2s^2\ 2p^3$ and for carbon is $1s^2\ 2s^2\ 2p^2$. This means that hydrogen has one valence electron which is unpaired, carbon has four valence electrons, all of which are unpaired, and nitrogen has five valence electrons, three of which are unpaired.

Step 2 : Arrange the electrons in the HCN molecule so that the outermost energy level in each atom is full.

The HCN molecule is represented below. Notice the three electron pairs between the nitrogen and carbon atom. Because these three covalent bonds are between the same two atoms, this is a *triple* bond.

Worked Example 10: Lewis notation: Atoms with variable valencies

Question: Represent the molecule H_2S using Lewis notation

Answer

Step 1 : Determine the number of valence electrons for each atom.

Hydrogen has an electron configuration of $1s^1$ and sulfur has an electron configuration of $1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^4$. Each hydrogen atom has one valence electron which is unpaired, and sulfur has six valence electrons. Although sulfur has a variable valency, we know that the sulfur will be able to form 2 bonds with the hydrogen atoms. In this case, the valency of sulfur must be two.

Step 2 : Arrange the atoms in the molecule so that the outermost energy level in each atom is full.

The H_2S molecule is represented below.

Another way of representing molecules is using **Couper notation**. In this case, only the electrons that are involved in the bond between the atoms are shown. A line is used for each covalent bond. Using Couper notation, a molecule of water and a molecule of HCN would be represented as shown in figures 4.6 and 4.7 below.

Figure 4.6: A water molecule represented using Couper notation

Figure 4.7: A molecule of HCN represented using Couper notation

Extension: Dative covalent bonds

A **dative covalent bond** (also known as a coordinate covalent bond) is a description of covalent bonding between two atoms in which both electrons shared in the bond come from the same atom. This happens when a Lewis base (an electron donor) donates a pair of electrons to a Lewis acid (an electron acceptor). Lewis acids and bases will be discussed in section 15.1 in chapter 15.

One example of a molecule that contains a dative covalent bond is the ammonium ion (NH_4^+) shown in the figure below. The hydrogen ion H^+ does not contain any electrons, and therefore the electrons that are in the bond that forms between this ion and the nitrogen atom, come only from the nitrogen.

Exercise: Atomic bonding and Lewis notation

1. Represent each of the following *atoms* using Lewis notation:
   (a) beryllium
   (b) calcium
   (c) lithium
2. Represent each of the following *molecules* using Lewis notation:
   (a) bromine gas (Br_2)

(b) calcium chloride ($CaCl_2$)
(c) carbon dioxide (CO_2)

3. Which of the three molecules listed above contains a double bond?
4. Two chemical reactions are described below.
   - nitrogen and hydrogen react to form ammonia
   - carbon and hydrogen bond to form a molecule of methane (CH_4)

   For each reaction, give:

   (a) the valency of each of the atoms involved in the reaction
   (b) the Lewis structure of the product that is formed
   (c) the chemical formula of the product
   (d) the name of the product
5. A chemical compound has the following Lewis notation:

   (a) How many valence electrons does element Y have?
   (b) What is the valency of element Y?
   (c) What is the valency of element X?
   (d) How many covalent bonds are in the molecule?
   (e) Suggest a name for the elements X and Y.

4.6 Electronegativity

Electronegativity is a measure of how strongly an atom pulls a shared electron pair towards it. The table below shows the electronegativities of a number of elements:

Table 4.1: Table of electronegativities for selected elements

Element	Electronegativity
Hydrogen (H)	2.1
Sodium (Na)	0.9
Magnesium (Mg)	1.2
Calcium (Ca)	1.0
Chlorine (Cl)	3.0
Bromine (Br)	2.8

Definition: Electronegativity
Electronegativity is a chemical property which describes the power of an atom to attract electrons towards itself.

The greater the electronegativity of an element, the stronger its attractive pull on electrons. For example, in a molecule of hydrogen bromide (HBr), the electronegativity of bromine (2.8) is higher than that of hydrogen (2.1), and so the shared electrons will spend more of their time closer to the bromine atom. Bromine will have a slightly negative charge, and hydrogen will have a slightly positive charge. In a molecule like hydrogen (H_2) where the electronegativities of the atoms in the molecule are the same, both atoms have a neutral charge.

Interesting Fact

The concept of electronegativity was introduced by *Linus Pauling* in 1932, and this became very useful in predicting the nature of bonds between atoms in molecules. In 1939, he published a book called 'The Nature of the Chemical Bond', which became one of the most influential chemistry books ever published. For this work, Pauling was awarded the Nobel Prize in Chemistry in 1954. He also received the Nobel Peace Prize in 1962 for his campaign against above-ground nuclear testing.

4.6.1 Non-polar and polar covalent bonds

Electronegativity can be used to explain the difference between two types of covalent bonds. **Non-polar covalent bonds** occur between two identical non-metal atoms, e.g. H_2, Cl_2 and O_2. Because the two atoms have the same electronegativity, the electron pair in the covalent bond is shared equally between them. However, if two different non-metal atoms bond then the shared electron pair will be pulled more strongly by the atom with the highest electronegativity. As a result, a **polar covalent bond** is formed where one atom will have a slightly negative charge and the other a slightly positive charge. This is represented using the symbols δ^+ (slightly positive) and δ^- (slightly negative). So, in a molecule such as hydrogen chloride (HCl), hydrogen is $H^{\delta+}$ and chlorine is $Cl^{\delta-}$.

4.6.2 Polar molecules

Some molecules with polar covalent bonds are **polar molecules**, e.g. H_2O. But not *all* molecules with polar covalent bonds are polar. An example is CO_2. Although CO_2 has two polar covalent bonds (between C^+ atom and the two O^- atoms), the molecule itself is not polar. The reason is that CO_2 is a linear molecule and is therefore symmetrical. So there is no difference in charge between the two ends of the molecule. The *polarity* of molecules affects properties such as *solubility*, *melting points* and *boiling points*.

Definition: Polar and non-polar molecules

A **polar molecule** is one that has one end with a slightly positive charge, and one end with a slightly negative charge. A **non-polar molecule** is one where the charge is equally spread across the molecule.

Exercise: Electronegativity

1. In a molecule of hydrogen chloride (HCl),
   (a) What is the electronegativity of hydrogen
   (b) What is the electronegativity of chlorine?
   (c) Which atom will have a slightly positive charge and which will have a slightly negative charge in the molecule?
   (d) Is the bond a non-polar or polar covalent bond?
   (e) Is the molecule polar or non-polar?

2. Complete the table below:

Molecule	Difference in electronegativity between atoms	Non-polar/polar covalent bond	Polar/non-polar molecule
H_2O			
HBr			
NO_2			
F_2			
CH_4			

4.7 Ionic Bonding

4.7.1 The nature of the ionic bond

You will remember that when atoms bond, electrons are either *shared* or they are *transferred* between the atoms that are bonding. In covalent bonding, electrons are shared between the atoms. There is another type of bonding, where electrons are *transferred* from one atom to another. This is called **ionic bonding**.

Ionic bonding takes place when the difference in electronegativity between the two atoms is more than 1.7. This usually happens when a metal atom bonds with a non-metal atom. When the difference in electronegativity is large, one atom will attract the shared electron pair much more strongly than the other, causing electrons to be transferred from one atom to the other.

Definition: Ionic bond
An ionic bond is a type of chemical bond based on the electrostatic forces between two oppositely-charged ions. When ionic bonds form, a metal donates an electron, due to a low electronegativity, to form a positive ion or cation. The non-metal atom has a high electronegativity, and therefore readily gains electrons to form negative ions or anions. The two or more ions are then attracted to each other by electrostatic forces.

Example 1:

In the case of NaCl, the difference in electronegativity is 2.1. Sodium has only one valence electron, while chlorine has seven. Because the electronegativity of chlorine is higher than the electronegativity of sodium, chlorine will attract the valence electron in the sodium atom very strongly. This electron from sodium is transferred to chlorine. Sodium has lost an electron and forms a Na^+ ion. Chlorine gains an electron and forms a Cl^- ion. The attractive force between the positive and negative ion is what holds the molecule together.

The balanced equation for the reaction is:

$$Na + Cl \rightarrow NaCl$$

This can be represented using Lewis notation:

Example 2:

Another example of ionic bonding takes place between magnesium (Mg) and oxygen (O) to form magnesium oxide (MgO). Magnesium has two valence electrons and an electronegativity of 1.2, while oxygen has six valence electrons and an electronegativity of 3.5. Since oxygen has

Figure 4.8: Ionic bonding in sodium chloride

a higher electronegativity, it attracts the two valence electrons from the magnesium atom and these electrons are transferred from the magnesium atom to the oxygen atom. Magnesium loses two electrons to form Mg^{2+}, and oxygen gains two electrons to form O^{2-}. The attractive force between the oppositely charged ions is what holds the molecule together.

The balanced equation for the reaction is:

$$2Mg + O_2 \rightarrow 2MgO$$

Because oxygen is a diatomic molecule, two magnesium atoms will be needed to combine with two oxygen atoms to produce two molecules of magnesium oxide (MgO).

Figure 4.9: Ionic bonding in magnesium oxide

Important: Notice that the number of electrons that is either lost or gained by an atom during ionic bonding, is the same as the **valency** of that element

---

Exercise: Ionic compounds

1. Explain the difference between a *covalent* and an *ionic* bond.
2. Magnesium and chlorine react to form magnesium chloride.
   (a) What is the difference in electronegativity between these two elements?
   (b) Give the chemical formula for:
      - a magnesium ion
      - a choride ion
      - the ionic compound that is produced during this reaction
   (c) Write a balanced chemical equation for the reaction that takes place.
3. Draw Lewis diagrams to represent the following ionic compounds:
   (a) sodium iodide (NaI)
   (b) calcium bromide ($CaBr_2$)
   (c) potassium chloride (KCl)

4.7.2 The crystal lattice structure of ionic compounds

Ionic substances are actually a combination of lots of ions bonded together into a giant molecule. The arrangement of ions in a regular, geometric structure is called a **crystal lattice**. So in fact NaCl does not contain one Na and one Cl ion, but rather a lot of these two ions arranged in a crystal lattice where the ratio of Na to Cl ions is 1:1. The structure of a crystal lattice is shown in figure 4.10.

Figure 4.10: The crystal lattice arrangement in an ionic compound (e.g. NaCl)

4.7.3 Properties of Ionic Compounds

Ionic compounds have a number of properties:

- Ions are arranged in a lattice structure
- Ionic solids are crystalline at room temperature
- The ionic bond is a strong electrical attraction. This means that ionic compounds are often hard and have high melting and boiling points
- Ionic compounds are brittle, and bonds are broken along planes when the compound is stressed
- Solid crystals don't conduct electricity, but ionic solutions do

4.8 Metallic bonds

4.8.1 The nature of the metallic bond

The structure of a metallic bond is quite different from covalent and ionic bonds. In a metal bond, the valence electrons are *delocalised*, meaning that an atom's electrons do not stay around that one nucleus. In a metallic bond, the positive atomic nuclei (sometimes called the 'atomic kernels') are surrounded by a sea of delocalised electrons which are attracted to the nuclei (figure 4.11).

Definition: Metallic bond
Metallic bonding is the electrostatic attraction between the positively charged atomic nuclei of metal atoms and the delocalised electrons in the metal.

Figure 4.11: Positive atomic nuclei (+) surrounded by delocalised electrons (•)

4.8.2 The properties of metals

Metals have several unique properties as a result of this arrangement:

- *Thermal conductors*

  Metals are good conductors of heat and are therefore used in cooking utensils such as pots and pans. Because the electrons are loosely bound and are able to move, they can transport heat energy from one part of the material to another.

- *Electrical conductors*

  Metals are good conductors of electricity, and are therefore used in electrical conducting wires. The loosely bound electrons are able to move easily and to transfer charge from one part of the material to another.

- *Shiny metallic lustre*

  Metals have a characteristic shiny appearance and are often used to make jewellery. The loosely bound electrons are able to absorb and reflect light at all frequencies, making metals look polished and shiny.

- *Malleable and ductile*

  This means that they can be bent into shape without breaking (malleable) and can be stretched into thin wires (ductile) such as copper, which can then be used to conduct electricity. Because the bonds are not fixed in a particular direction, atoms can slide easily over one another, making metals easy to shape, mould or draw into threads.

- *Melting point*

  Metals usually have a high melting point and can therefore be used to make cooking pots and other equipment that needs to become very hot, without being damaged. The high melting point is due to the high strength of metallic bonds.

- *Density*

  Metals have a high density because their atoms are packed closely together.

?

Exercise: Chemical bonding

1. Give two examples of everyday objects that contain..
   (a) covalent bonds
   (b) ionic bonds

(c) metallic bonds

2. Complete the table which compares the different types of bonding:

	Covalent	**Ionic**	**Metallic**
Types of atoms involved			
Nature of bond between atoms			
Melting Point (high/low)			
Conducts electricity? (yes/no)			
Other properties			

3. Complete the table below by identifying the type of bond (covalent, ionic or metallic) in each of the compounds:

Molecular formula	**Type of bond**
H_2SO_4	
FeS	
NaI	
$MgCl_2$	
Zn	

4. Which of these substances will conduct electricity most effectively? Give a reason for your answer.
5. Use your knowledge of the different types of bonding to explain the following statements:
   (a) Swimming during an electric thunderstorm (i.e. where there is lightning) can be very dangerous.
   (b) Most jewellery items are made from metals.
   (c) Plastics are good insulators.

4.9 Writing chemical formulae

4.9.1 The formulae of covalent compounds

To work out the formulae of covalent compounds, we need to use the valency of the atoms in the compound. This is because the valency tells us how many bonds each atom can form. This in turn can help to work out how many atoms of each element are in the compound, and therefore what its formula is. The following are some examples where this information is used to write the chemical formula of a compound.

Worked Example 11: Formulae of covalent compounds

Question: Write the chemical formula for water

Answer
Step 1 : Write down the elements that make up the compound.
A molecule of water contains the elements *hydrogen* and *oxygen*.

Step 2 : Determine the valency of each element
The valency of hydrogen is 1 and the valency of oxygen is 2. This means that oxygen can form two bonds with other elements and each of the hydrogen atoms can form one.

Step 3 : Write the chemical formula
Using the valencies of hydrogen and oxygen, we know that in a single water molecule, two hydrogen atoms will combine with one oxygen atom. The chemical formula for water is therefore:

$$\mathbf{H_2O}.$$

Worked Example 12: Formulae of covalent compounds

Question: Write the chemical formula for magnesium oxide

Answer
Step 1 : Write down the elements that make up the compound.
A molecule of magnesium oxide contains the elements *magnesium* and *oxygen*.

Step 2 : Determine the valency of each element
The valency of magnesium is 2, while the valency of oxygen is also 2. In a molecule of magnesium oxide, one atom of magnesium will combine with one atom of oxygen.

Step 3 : Write the chemical formula
The chemical formula for magnesium oxide is therefore:

$$\mathbf{MgO}$$

Worked Example 13: Formulae of covalent compounds

Question: Write the chemical formula for copper (II) chloride.

Answer
Step 1 : Write down the elements that make up the compound.
A molecule of copper (II) chloride contains the elements *copper* and *chlorine*.

Step 2 : Determine the valency of each element
The valency of copper is 2, while the valency of chlorine is 1. In a molecule of copper (II) chloride, two atoms of chlorine will combine with one atom of copper.

Step 3 : Write the chemical formula
The chemical formula for copper (II) chloride is therefore:

$$\mathbf{CuCl_2}$$

4.9.2 The formulae of ionic compounds

The overall charge of an ionic compound will always be zero and so the negative and positive charge must be the same size. We can use this information to work out what the chemical formula of an ionic compound is if we know the charge on the individual ions. In the case of NaCl for example, the charge on the sodium is +1 and the charge on the chlorine is -1. The charges balance (+1-1=0) and therefore the ionic compound is neutral. In MgO, magnesium has a charge of +2 and oxygen has a charge of -2. Again, the charges balance and the compound is neutral. Positive ions are called **cations** and negative ions are called **anions**.

Some ions are made up of groups of atoms, and these are called **compound ions**. It is a good idea to learn the compound ions that are shown in table 4.2

Table 4.2: Table showing common compound ions and their formulae

Name of compound ion	formula
Carbonate	CO_3^{2-}
sulphate	SO_4^{2-}
Hydroxide	OH^-
Ammonium	NH_4^+
Nitrate	NO_3^-
Hydrogen carbonate	HCO_3^-
Phosphate	PO_4^{3-}
Chlorate	ClO_3^-
Cyanide	CN^-
Chromate	CrO_4^{2-}
Permanganate	MnO_4^-

In the case of ionic compounds, the valency of an ion is the same as its charge (Note: valency is always expressed as a *positive* number e.g. valency of the chloride ion is 1 and not -1). Since an ionic compound is always *neutral*, the positive charges in the compound must balance out the negative. The following are some examples:

Worked Example 14: Formulae of ionic compounds

Question: Write the chemical formula for potassium iodide.

Answer
Step 1 : Write down the ions that make up the compound.
Potassium iodide contains potassium and iodide ions.

Step 2 : Determine the valency and charge of each ion.
Potassium iodide contains the ions K^+ (valency = 1; charge = +1) and I^- (valency = 1; charge = -1). In order to balance the charge in a single molecule, one atom of potassium will be needed for every one atom of iodine.

Step 3 : Write the chemical formula
The chemical formula for potassium iodide is therefore:

KI

Worked Example 15: Formulae of ionic compounds

Question: Write the chemical formula for sodium sulphate.

Answer
Step 1 : Write down the ions that make up the compound.
Sodium sulphate contains sodium ions and sulphate ions.

Step 2 : Determine the valency and charge of each ion.
Na^+ (valency = 1; charge = +1) and $SO_4{}^{2-}$ (valency = 2; charge = -2).

Step 3 : Write the chemical formula.
Two sodium ions will be needed to balance the charge of the sulphate ion. The chemical formula for sodium sulphate is therefore:

$$\mathbf{Na_2SO_4{}^{2-}}$$

Worked Example 16: Formulae of ionic compounds

Question: Write the chemical formula for calcium hydroxide.

Answer
Step 1 : Write down the ions that make up the compound.
Calcium hydroxide contains calcium ions and hydroxide ions.

Step 2 : Determine the valency and charge of each ion.
Calcium hydroxide contains the ions Ca^{2+} (charge = +2) and OH^- (charge = -1). In order to balance the charge in a single molecule, two hydroxide ions will be needed for every ion of calcium.

Step 3 : Write the chemical formula.
The chemical formula for calcium hydroxide is therefore:

$$\mathbf{Ca(OH)_2}$$

Exercise: Chemical formulae

1. Copy and complete the table below:

Compound	Cation	Anion	Formula
	Na^+	Cl^-	
potassium bromide		Br^-	
	NH_4^+	Cl^-	
potassium chromate			
			PbI
potassium permanganate			
calcium phosphate			

2. Write the chemical formula for each of the following compounds:
   (a) hydrogen cyanide
   (b) carbon dioxide
   (c) sodium carbonate
   (d) ammonium hydroxide
   (e) barium sulphate
   (f) potassium permanganate

4.10 The Shape of Molecules

4.10.1 Valence Shell Electron Pair Repulsion (VSEPR) theory

The shape of a covalent molecule can be predicted using the Valence Shell Electron Pair Repulsion (VSEPR) theory. This is a model in chemistry that tries to predict the shapes of molecules. Very simply, VSEPR theory says that the electron pairs in a molecule will arrange themselves around the central atom of the molecule so that the repulsion between their negative charges is as small as possible. In other words, the electron pairs arrange themselves so that they are as far apart as they can be. Depending on the number of electron pairs in the molecule, it will have a different shape.

Definition: Valence Shell Electron Pair Repulsion Theory

Valence shell electron pair repulsion theory (VSEPR) is a model in chemistry, which is used to predict the shape of individual molecules, based upon the extent of their electron-pair repulsion.

VSEPR theory is based on the idea that the geometry of a molecule is mostly determined by repulsion among the pairs of electrons around a central atom. The pairs of electrons may be bonding or non-bonding (also called lone pairs). Only valence electrons of the central atom influence the molecular shape in a meaningful way.

4.10.2 Determining the shape of a molecule

To predict the shape of a covalent molecule, follow these steps:

Step 1:

Draw the molecule using Lewis notation. Make sure that you draw *all* the electrons around the molecule's central atom.

Step 2:

Count the number of electron pairs around the central atom.

Step 3:

Determine the basic geometry of the molecule using the table below. For example, a molecule with two electron pairs around the central atom has a *linear* shape, and one with four electron pairs around the central atom would have a *tetrahedral* shape. The situation is actually more

Table 4.3: The effect of electron pairs in determining the shape of molecules

Number of electron pairs	Geometry
2	linear
3	trigonal planar
4	tetrahedral
5	trigonal bipyramidal
6	octahedral

complicated than this, but this will be discussed later in this section.

Figure 4.12 shows each of these shapes. Remember that the shapes are 3-dimensional, and so you need to try to imagine them in this way. In the diagrams, the shaded part represents those parts of the molecule that are 'in front', while the dashed lines represent those parts that are 'at the back' of the molecule.

Figure 4.12: Some common molecular shapes

Worked Example 17: The shapes of molecules

Question: Determine the shape of a molecule of O_2

Answer
Step 1 : Draw the molecule using Lewis notation

Step 2 : Count the number of electron pairs around the central atom
There are two electron pairs.

Step 3 : Determine the basic geometry of the molecule
Since there are two electron pairs, the molecule must be linear.

Worked Example 18: The shapes of molecules

Question: Determine the shape of a molecule of BF_3

Answer
Step 1 : Draw the molecule using Lewis notation

Step 2 : Count the number of electron pairs around the central atom
There are three electron pairs.

Step 3 : Determine the basic geometry of the molecule
Since there are three electron pairs, the molecule must be trigonal planar.

Extension: More about molecular shapes

Determining the shape of a molecule can be a bit more complicated. In the examples we have used above, we looked only at the number of **bonding electron pairs** when we were trying to decide on the molecules' shape. But there are also other electron pairs in the molecules. These electrons, which are not involved in bonding but which are also around the central atom, are called **lone pairs**. The worked example below will give you an indea of how these lone pairs can affect the shape of the molecule.

Worked Example 19: Advanced

Question: Determine the shape of a molecule of NH_3

Answer
Step 1 : Draw the molecule using Lewis notation

Step 2 : Count the number of electron pairs around the central atom
There are four electron pairs.

Step 3 : Determine the basic geometry of the molecule
Since there are four electron pairs, the molecule must be tetrahedral.

Step 4 : Determine how many lone pairs are around the central atom
There is one lone pair of electrons and this will affect the shape of the molecule.

Step 5 : Determine the final shape of the molecule
The lone pair needs more space than the bonding pairs, and therefore pushes the three hydrogen atoms together a little more. The bond angles between the hydrogen and nitrogen atoms in the molecule become 106 degrees, rather than the usual 109 degrees of a tetrahedral molecule. The shape of the molecule is *trigonal pyramidal*.

Activity :: Group work : Building molecular models

In groups, you are going to build a number of molecules using marshmallows to represent the atoms in the molecule, and toothpicks to represent the bonds between the atoms. In other words, the toothpicks will hold the atoms (marshmallows) in the molecule together. Try to use different coloured marshmallows to represent different elements.

You will build models of the following molecules:
HCl, CH_4, H_2O, HBr and NH_3
For each molecule, you need to:

- Determine the basic geometry of the molecule
- Build your model so that the atoms are as far apart from each other as possible (remember that the electrons around the central atom will try to avoid the repulsions between them).
- Decide whether this shape is accurate for that molecule or whether there are any lone pairs that may influence it.
- Adjust the position of the atoms so that the bonding pairs are further away from the lone pairs.
- How has the shape of the molecule changed?
- Draw a simple diagram to show the shape of the molecule. It doesn't matter if it is not 100% accurate. This exercise is only to help you to visualise the 3-dimensional shapes of molecules.

Do the models help you to have a clearer picture of what the molecules look like? Try to build some more models for other molecules you can think of.

4.11 Oxidation numbers

When reactions occur, an exchange of electrons takes place. **Oxidation** is the *loss* of electrons from an atom, while **reduction** is the *gain* of electrons by an atom. By giving elements an oxidation number, it is possible to keep track of whether that element is losing or gaining electrons during a chemical reaction. The loss of electrons in one part of the reaction, must be balanced by a gain of electrons in another part of the reaction.

Definition: Oxidation number
A simplified way of understanding an oxidation number is to say that it is the charge an atom would have if it was in a compound composed of ions.

There are a number of rules that you need to know about oxidation numbers, and these are listed below. These will probably not make much sense at first, but once you have worked through some examples, you will soon start to understand!

1. **Rule 1:** An element always has an oxidation number of zero, since it is neutral.

   In the reaction $H_2 + Br_2 \rightarrow 2HBr$, the oxidation numbers of hydrogen and bromine on the left hand side of the equation are both zero.

2. **Rule 2:** In most cases, an atom that is part of a molecule will have an oxidation number that has the same numerical value as its valency.

3. **Rule 3:** Monatomic ions have an oxidation number that is equal to the charge on the ion.

   The chloride ion Cl^- has an oxidation number of -1, and the magnesium ion Mg^{2+} has an oxidation number of $+2$.

4. **Rule 4:** In a molecule, the oxidation number for the whole molecule will be zero, unless the molecule has a charge, in which case the oxidation number is equal to the charge.

5. **Rule 5:** Use a table of electronegativities to determine whether an atom has a positive or a negative oxidation number. For example, in a molecule of water, oxygen has a higher electronegativity so it must be negative because it attracts electrons more strongly. It will have a negative oxidation number (-2). Hydrogen will have a positive oxidation number ($+1$).

6. **Rule 6:** An oxygen atom usually has an oxidation number of -2, although there are some cases where its oxidation number is -1.

7. **Rule 7:** The oxidation number of hydrogen is usually $+1$. There are some exceptions where its oxidation number is -1.

8. **Rule 8:** In most compounds, the oxidation number of the halogens is -1.

Important: You will notice that the oxidation number of an atom is the same as its valency. Whether an oxidation number os positive or negative, is determined by the electronegativities of the atoms involved.

Worked Example 20: Oxidation numbers

Question: Give the oxidation numbers for all the atoms in the reaction between sodium and chlorine to form sodium chloride.

$$Na + Cl \rightarrow NaCl$$

Answer

Step 1 : Determine which atom will have a positive or negative oxidation number

Sodium will have a positive oxidation number and chlorine will have a negative oxidation number.

Step 2 : Determine the oxidation number for each atom

Sodium (group 1) will have an oxidation number of $+1$. Chlorine (group 7) will have an oxidation number of -1.

Step 3 : Check whether the oxidation numbers add up to the charge on the molecule

In the equation $Na + Cl \rightarrow NaCl$, the overall charge on the NaCl molecule is $+1$-1=0. This is correct since NaCl is neutral. This means that, in a molecule of NaCl, sodium has an oxidation number of $+1$ and chlorine has an oxidation number of -1.

Worked Example 21: Oxidation numbers

Question: Give the oxidation numbers for all the atoms in the reaction between hydrogen and oxygen to produce water. The unbalanced equation is shown below:

$$H_2 + O_2 \rightarrow H_2O$$

Answer

Step 1 : Determine which atom will have a positive or negative oxidation number

Hydrogen will have a positive oxidation number and oxygen will have a negative oxidation number.

Step 2 : Determine the oxidation number for each atom

Hydrogen (group 1) will have an oxidation number of $+1$. Oxygen (group 6) will have an oxidation number of -2.

Step 3 : Check whether the oxidation numbers add up to the charge on the molecule

In the reaction $H_2 + O_2 \rightarrow H_2O$, the oxidation numbers for hydrogen and oxygen (on the left hand side of the equation) are zero since these are elements. In the water molecule, the sum of the oxidation numbers is 2(+1)-2=0. This is correct since the oxidation number of water is zero. Therefore, in water, hydrogen has an oxidation number of $+1$ and oxygen has an oxidation number of -2.

Worked Example 22: Oxidation numbers

Question: Give the oxidation number of sulfur in a sulphate (SO_4^{2-}) ion

Answer

Step 1 : Determine which atom will have a positive or negative oxidation number

Sulfur has a positive oxidation number and oxygen will have a negative oxidation number.

Step 2 : Determine the oxidation number for each atom

Oxygen (group 6) will have an oxidation number of -2. The oxidation number of sulfur at this stage is uncertain.

Step 3 : Determine the oxidation number of sulfur by using the fact that the oxidation numbers of the atoms must add up to the charge on the molecule

In the polyatomic SO_4^{2-} ion, the sum of the oxidation numbers must be -2. Since there are four oxygen atoms in the ion, the total charge of the oxygen is -8. If the overall charge of the ion is -2, then the oxidation number of sulfur must be $+6$.

Exercise: Oxidation numbers

1. Give the oxidation numbers for each element in the following chemical compounds:
   (a) NO_2
   (b) $BaCl_2$
   (c) H_2SO_4
2. Give the oxidation numbers for the reactants and products in each of the following reactions:
   (a) $C + O_2 \rightarrow CO_2$
   (b) $N_2 + 3H_2 \rightarrow 2NH_3$
   (c) Magnesium metal burns in oxygen

4.12 Summary

- A **chemical bond** is the physical process that causes atoms and molecules to be attracted together and to be bound in new compounds.
- Atoms are more **reactive**, and therefore more likely to bond, when their outer electron orbitals are not full. Atoms are less reactive when these outer orbitals contain the maximum number of electrons. This explains why the noble gases do not combine to form molecules.
- There are a number of **forces** that act between atoms: attractive forces between the positive nucleus of one atom and the negative electrons of another; repulsive forces between like-charged electrons, and repulsion between like-charged nuclei.
- Chemical bonding occurs when the **energy** of the system is at its lowest.
- **Bond length** is the distance between the nuclei of the atoms when they bond.
- **Bond energy** is the energy that must be added to the system for the bonds to break.
- When atoms bond, electrons are either shared or exchanged.
- **Covalent bonding** occurs between the atoms of non-metals and involves a sharing of electrons so that the orbitals of the outermost energy levels in the atoms are filled.
- The **valency** of an atom is the number of bonds that it can form with other atoms.
- A **double** or **triple bond** occurs if there are two or three electron pairs that are shared between the same two atoms.
- A **dative covalent bond** is a bond between two atoms in which both the electrons that are shared in the bond come from the same atom.
- **Lewis** and **Couper** notation are two ways of representing molecular structure. In Lewis notation, dots and crosses are used to represent the valence electrons around the central atom. In Couper notation, lines are used to represent the bonds between atoms.
- **Electronegativity** measures how strongly an atom draws electrons to it.
- Electronegativity can be used to explain the difference between two types of covalent bonds: **polar covalent bonds** (between non-identical atoms) and **non-polar covalent bonds** (between identical atoms).
- An **ionic bond** occurs between atoms where the difference in electronegativity is greater than 2.1. An exchange of electrons takes place and the atoms are held together by the electrostatic force of attraction between oppositely-charged ions.
- Ionic solids are arranged in a **crystal lattice** structure.

- Ionic compounds have a number of specific **properties**, including their high melting and boiling points, brittle nature, the arrangement of solids in a lattice structure and the ability of ionic solutions to conduct electricity.
- A **metallic bond** is the electrostatic attraction between the positively charged nuclei of metal atoms and the delocalised electrons in the metal.
- Metals also have a number of properties, including their ability to conduct heat and electricity, their metallic lustre, the fact that they are both malleable and ductile, and their high melting point and density.
- The valency of atoms, and the way they bond, can be used to determine the **chemical formulae** of compounds.
- The **shape of molecules** can be predicted using the VSEPR theory, which uses the arrangement of electrons around the central atom to determine the most likely shape of the molecule.
- **Oxidation numbers** are used to determine whether an atom has gained or lost electrons during a chemical reaction.

Exercise: Summary exercise

1. Give **one word/term** for each of the following descriptions.
   (a) The distance between two atoms in a molecule
   (b) A type of chemical bond that involves the transfer of electrons from one atom to another.
   (c) A measure of an atom's ability to attract electrons to it.
2. Which ONE of the following best describes the bond formed between an H^+ ion and the NH_3 molecule?
   (a) Covalent bond
   (b) Dative covalent (coordinate covalent) bond
   (c) Ionic Bond
   (d) Hydrogen Bond
3. Explain the meaning of each of the following terms:
   (a) valency
   (b) bond energy
   (c) covalent bond
4. Which of the following reactions will *not* take place? Explain your answer.
   (a) $\mathrm{H} + \mathrm{H} \rightarrow \mathrm{H_2}$
   (b) $\mathrm{Ne} + \mathrm{Ne} \rightarrow \mathrm{Ne_2}$
   (c) $\mathrm{I} + \mathrm{I} \rightarrow \mathrm{I_2}$
5. Draw the Lewis structure for each of the following:
   (a) calcium
   (b) iodine
   (c) hydrogen bromide (HBr)
   (d) nitrogen dioxide (NO_2)
6. Given the following Lewis structure, where X and Y each represent a different element...

   (a) What is the valency of X?

(b) What is the valency of Y?

(c) Which elements could X and Y represent?

7. A molecule of ethane has the formula C_2H_6. Which of the following diagrams (Couper notation) accurately represents this molecule?

(a)

```
      H   H
      |   |
H — C = C — H
      |   |
      H   H
```

(b)

```
      H
      |
H — C — C — H
      |   |
      H   H
```

(c)

```
      H   H
      |   |
H — C — C — H
      |   |
      H   H
```

8. Potassium dichromate is dissolved in water.

(a) Give the name and chemical formula for each of the ions in solution.

(b) What is the chemical formula for potassium dichromate?

(c) Give the oxidation number for each element in potassium dichromate.

Chapter 5

Intermolecular Forces - Grade 11

In the previous chapter, we discussed the different forces that exist *between atoms* (intramolecular forces). When atoms are joined to one another they form molecules, and these molecules in turn have forces that bind them together. These forces are known as **intermolecular forces**, and we are going to look at them in more detail in this next section.

Definition: Intermolecular forces
Intermolecular forces are forces that act between stable molecules.

You will also remember from the previous chapter, that we can describe molecules as being either **polar** or **non-polar**. A polar molecule is one in which there is a difference in electronegativity between the atoms in the molecule, such that the shared electron pair spends more time close to the atom that attracts it more strongly. The result is that one end of the molecule will have a slightly positive charge (δ^+), and the other end will have a slightly negative charge (δ^+). The molecule is said to be a **dipole**. However, it is important to remember that just because the bonds within a molecule are polar, the molecule itself may not necessarily be polar. The shape of the molecule may also affect its polarity. A few examples are shown in table 5.1 to refresh your memory!

5.1 Types of Intermolecular Forces

It is important to be able to recognise whether the molecules in a substance are polar or non-polar because this will determine what type of inermolecular forces there are. This is important in explaining the properties of the substance.

1. **Van der Waals forces**

   These intermolecular forces are named after a Dutch physicist called Johannes van der Waals (1837 -1923), who recognised that there were weak attractive and repulsive forces between the molecules of a gas, and that these forces caused gases to deviate from 'ideal gas' behaviour. Van der Waals forces are *weak* intermolecular forces, and can be divided into three types:

   (a) *Dipole-dipole forces*

   Figure 5.1 shows a simplified dipole molecule, with one end slightly positive and the other slightly negative.

   When one dipole molecule comes into contact with another dipole molecule, the positive pole of the one molecule will be attracted to the negative pole of the other, and the molecules will be held together in this way (figure 5.2). Examples of molecules that are held together by dipole-dipole forces are HCl, FeS, KBr, SO_2 and NO_2.

   (b) *Ion-dipole forces*

   As the name suggests, this type of intermolecular force exists between an ion and a dipole molecule. You will remember that an *ion* is a charged atom, and this will

Table 5.1: Polarity in molecules with different atomic bonds and molecular shapes

Molecule	**Chemical formula**	**Bond between atoms**	**Shape of molecule**	**Polarity of molecule**
Hydrogen	H_2	Covalent	Linear molecule **H** —— **H**	Non-polar
Hydrogen chloride	HCl	Polar covalent	Linear molecule $\mathbf{H}^{\delta+}$ —— $\mathbf{Cl}^{\delta-}$	Polar
Carbon tetrafluoromethane	CF_4	Polar covalent	Tetrahedral molecule $F^{\delta-}$ $F^{\delta-}$ —— $C^{\delta+}$ —— $F^{\delta-}$ $F^{\delta-}$	Non-polar

Figure 5.1: A simplified diagram of a dipole molecule

Figure 5.2: Two dipole molecules are held together by the attractive force between their oppositely charged poles

be attracted to one of the charged ends of the polar molecule. A positive ion will be attracted to the negative pole of the polar molecule, while a negative ion will be attracted to the positive pole of the polar molecule. This can be seen when sodium chloride (NaCl) dissolves in water. The positive sodium ion (Na^+) will be attracted to the slightly negative oxygen atoms in the water molecule, while the negative chloride ion (Cl^-) is attracted to the slightly positive hydrogen atom. These intermolecular forces weaken the ionic bonds between the sodium and chloride ions so that the sodium chloride dissolves in the water (figure 5.3).

(c) *London forces*

These intermolecular forces are also sometimes called 'dipole- induced dipole' or 'momentary dipole' forces. Not all molecules are polar, and yet we know that there are also intermolecular forces between non-polar molecules such as carbon dioxide. In

Figure 5.3: Ion-dipole forces in a sodium chloride solution

non-polar molecules the electronic charge is evenly distributed but it is possible that at a particular moment in time, the electrons might not be evenly distributed. The molecule will have a *temporary dipole*. In other words, each end of the molecules has a slight charge, either positive or negative. When this happens, molecules that are next to each other attract each other very weakly. These London forces are found in the halogens (e.g. F_2 and I_2), the noble gases (e.g. Ne and Ar) and in other non-polar molecules such as carbon dioxide and methane.

2. **Hydrogen bonds**

   As the name implies, this type of intermolecular bond involves a hydrogen atom. The hydrogen must be attached to another atom that is strongly electronegative, such as oxygen, nitrogen or fluorine. Water molecules for example, are held together by hydrogen bonds between the hydrogen atom of one molecule and the oxygen atom of another (figure 5.4). Hydrogen bonds are stronger than van der Waals forces. It is important to note however, that both van der Waals forces and hydrogen bonds are weaker than the covalent and ionic bonds that exist between *atoms*.

Figure 5.4: Two representations showing the hydrogen bonds between water molecules: space-filling model and structural formula.

Exercise: Types of intermolecular forces

1. Complete the following table by placing a tick to show which type of intermolecular force occurs in each substance:

Formula	Dipole-dipole	Momentary dipole	Ion-dipole	hydrogen bond
HCl				
CO_2				
I_2				
H_2O				
KI(aq)				
NH_3				

2. In which of the substances above are the intermolecular forces...
   (a) strongest
   (b) weakest

5.2 Understanding intermolecular forces

The types of intermolecular forces that occur in a substance will affect its properties, such as its **phase**, **melting point** and **boiling point**. You should remember, if you think back to the kinetic theory of matter, that the *phase* of a substance is determined by how strong the forces are between its particles. The weaker the forces, the more likely the substance is to exist as a gas because the particles are far apart. If the forces are very strong, the particles are held closely together in a solid structure. Remember also that the *temperature* of a material affects the energy of its particles. The more energy the particles have, the more likely they are to be able to overcome the forces that are holding them together. This can cause a change in phase.

Definition: Boiling point
The temperature at which a material will change from being a liquid to being a gas.

Definition: Melting point
The temperature at which a material will change from being a solid to being a liquid.

Now look at the data in table 5.2.

Formula	Formula mass	Melting point (^{0}C)	Boiling point (^{0}C) at 1 atm
He	4	-270	-269
Ne	20	-249	-246
Ar	40	-189	-186
F_2	38	-220	-188
Cl_2	71	-101	-35
Br_2	160	-7	58
NH_3	17	-78	-33
H_2O	18	0	100
HF	20	-83	20

Table 5.2: Melting point and boiling point of a number of chemical substances

The melting point and boiling point of a substance, give us information about the *phase* of the substance at room temperature, and the *strength of the intermolecular forces*. The examples below will help to explain this.

Example 1: Fluorine (F_2)

Phase at room temperature

Fluorine (F_2) has a melting point of -220^{0}C and a boiling point of -188^{0}C. This means that for any temperature that is greater than -188^{0}C, fluorine will be a gas. At temperatures below -220^{0}C, fluorine would be a solid, and at any temperature inbetween these two, fluorine will be a liquid. So, at room temperature, fluorine exists as a gas.

Strength of intermolecular forces

What does this information tell us about the intermolecular forces in fluorine? In fluorine, these forces must be very weak for it to exist as a gas at room temperature. Only at temperatures below -188^{0}C will the molecules have a low enough energy that they will come close enough to each other for forces of attraction to act between the molecules. The intermolecular forces in fluorine are very weak **van der Waals** forces because the molecules are *non-polar*.

Example 2: Hydrogen fluoride (HF)

Phase at room temperature

For temperatures below -83^{0}C, hydrogen fluoride is a solid. Between -83^{0}C and 20^{0}C, it exists as a liquid, and if the temperature is increased above 20^{0}C, it will become a gas.

Strength of intermolecular forces

What does this tell us about the intermolecular forces in hydrogen fluoride? The forces are stronger than those in fluorine, because more energy is needed for fluorine to change into the gaseous phase. In other words, more energy is needed for the intermolecular forces to be overcome so that the molecules can move further apart. Intermolecular forces will exist between the hydrogen atom of one molecule and the fluorine atom of another. These are **hydrogen bonds**, which are stronger than van der Waals forces.

What do you notice about water? Luckily for us, water behaves quite differently from the rest of the halides. Imagine if water were like ammonia (NH_3), which is a gas above a temperature of -33^{0}C! There would be no liquid water on the planet, and that would mean that no life would be able to survive here. The hydrogen bonds in water are particularly strong and this gives water unique qualities when compared to other molecules with hydrogen bonds. This will be discussed more in chapter **??** deals with this in more detail. You should also note that the strength of the intermolecular forces increases with an increase in formula mass. This can be seen by the increasing melting and boiling points of substances as formula mass increases.

?

Exercise: Applying your knowledge of intermolecular forces

Refer to the data in table 5.2 and then use your knowledge of different types of intermolecular forces to explain the following statements:

- The boiling point of F_2 is much lower than the boiling point of NH_3
- At room temperature, many elements exist naturally as gases
- The boiling point of HF is higher than the boiling point of Cl_2
- The boiling point of water is much higher than HF, even though they both contain hydrogen bonds

---

5.3 Intermolecular forces in liquids

Intermolecular forces affect a number of properties in liquids:

- **Surface tension**

  You may have noticed how some insects are able to walk across a water surface, and how leaves float in water. This is because of surface tension. In water, each molecule is held to the surrounding molecules by strong hydrogen bonds. Molecules in the centre of the liquid are completely surrounded by other molecules, so these forces are exerted in all directions. However, molecules at the surface do not have any water molecules above them to pull them upwards. Because they are only pulled sideways and downwards, the water molecules at the surface are held more closely together. This is called **surface tension**.

Figure 5.5: Surface tension in a liquid

- **Capillarity**

Activity :: Investigation : Capillarity

Half fill a beaker with water and hold a hollow glass tube in the centre as shown below. Mark the level of the water in the glass tube, and look carefully at the shape of the air-water interface in the tube. What do you notice?

At the air-water interface, you will notice a **meniscus**, where the water appears to dip in the centre. In the glass tube, the attractive forces between the glass and the water are stronger than the intermolecular forces between the water molecules. This causes the water to be held more closely to the glass, and a meniscus forms. The forces between the glass and the water also mean that the water can be 'pulled up' higher when it is in the tube than when it is in teh beaker. Capillarity is the surface tension that occurs in liquids that are inside tubes.

- **Evaporation**

  In a liquid, each particle has kinetic energy, but some particles will have more energy than others. We therefore refer to the *average* kinetic energy of the molecules when we describe the liquid. When the liquid is heated, those particles which have the highest energy will be able to overcome the intermolecular forces holding them in the liquid phase, and will become a gas. This is called **evaporation**. Evaporation occurs when a liquid changes to a gas. The stronger the intermolecular forces in a liquid, the higher the temperature of the molecules will have to be for it to become a gas. You should note that a liquid doesn't necessarily have to reach boiling point before evaporation can occur. Evaporation takes place all the time. You will see this if you leave a glass of water outside in the sun. Slowly the water level will drop over a period of time.

  What happens then to the molecules of water that remain in the liquid? Remember that it was the molecules with the highest energy that left the liquid. This means that the average kinetic energy of the remaining molecules will decrease, and so will the *temperature* of the liquid.

  A similar process takes place when a person sweats during exercise. When you exercise, your body temperature increases and you begin to release moisture (sweat) through the pores in your skin. The sweat quickly evaporates and causes the temperature of your skin to drop. This helps to keep your body temperature at a level that is suitable forit to function properly.

Interesting Fact

Transpiration in plants - Did you know that plants also 'sweat'? In plants, this is called *transpiration*, and a plant will lose water through spaces in the leaf surface called *stomata*. Although this water loss is important in the survival of a plant, if a plant loses too much water, it will die. Plants that live in very hot, dry places such as deserts, must be specially adapted to reduce the amount of water that transpires (evaporates) from their leaf surface. Desert plants have some amazing adaptations to deal with this problem! Some have hairs on their leaves, which reflect sunlight so that the temperature is not as high as it would be, while others have a thin waxy layer covering their leaves, which reduces water loss. Some plants are even able to close their stomata during the day when temperatures (and therefore transpiration) are highest.

Important: In the same way that intermolecular forces affect the properties of liquids, they also affect the properties of solids. For example, the stronger the intermolecular forces between the particles that make up the solid, the *harder* the solid is likely to be, and the higher its *melting point* is likely to be.

5.4 Summary

- **Intermolecular forces** are the forces that act between stable molecules.
- The **type** of intermolecular force in a substance, will depend on the **nature of the molecules**.
- **Polar molecules** have an unequal distribution of charge, meaning that one part of the molecule is slightly positive and the other part is slightly negative. **Non-polar molecules** have an equal distribution of charge.

- There are three types of **Van der Waal's forces**. These are dipole-dipole, ion-dipole and London forces (momentary dipole).
- **Dipole-dipole** forces exist between two **polar molecules**, for example between two molecules of hydrogen chloride.
- **Ion-dipole** forces exist between **ions and dipole molecules**. The ion is attracted to the part of the molecule that has an opposite charge to its own. An example of this is when an ionic solid such as sodium chloride dissolves in water.
- **Momentary dipole** forces occur between two **non-polar molecules**, where at some point there is an uequal distribution of charge in the molecule. For example, there are London forces between two molecules of carbon dioxide.
- **Hydrogen bonds** occur between **hydrogen atoms** and other **atoms that have a high electronegativity** such as oxygen, nitrogen and fluorine. The hydrogen atom in one molecule will be attracted to the nitrogen atom in another molecule, for example. There are hydrogen bonds between water molecules and between ammonia molecules.
- Intermolecular forces affect the **properties** of substances. For example, the stronger the intermolecular forces, the higher the **melting point** of that substance, and the more likely that substance is to exist as a solid or liquid. Its **boiling point** will also be higher.
- In **liquids**, properties such as **surface tension**, **capillarity** and **evaporation** are the result of intermolecular forces.

Exercise: Summary Exercise

1. Give one word or term for each of the following descriptions:
   (a) The tendency of an atom in a molecule to attract bonding electrons.
   (b) A molecule that has an unequal distribution of charge.
   (c) A charged atom.
2. For each of the following questions, choose the one correct answer from the list provided.
   (a) The following table gives the melting points of various hydrides:

Hydride	**Melting point** (^{0}C)
HI	-50
NH_3	-78
H_2S	-83
CH_4	-184

   In which of these hydrides does hydrogen bonding occur?
   i. HI only
   ii. NH_3 only
   iii. HI and NH_3 only
   iv. HI, NH_3 and H_2S
   (IEB Paper 2, 2003)

   (b) Refer to the list of substances below:
   HCl, Cl_2, H_2O, NH_3, N_2, HF
   Select the true statement from the list below:
   i. NH_3 is a non-polar molecule
   ii. The melting point of NH_3 will be higher than for Cl_2
   iii. Ion-dipole forces exist between molecules of HF
   iv. At room temperature N_2 is usually a liquid

3. The respective boiling points for four chemical substances are given below:
   Hydrogen sulphide -60^{0}C
   Ammonia -33^{0}C
   Hydrogen fluoride 20^{0}C
   Water 100^{0}C
   (a) Which one of the substances exhibits the strongest forces of attraction between its molecules in the liquid state?
   (b) Give the name of the force responsible for the relatively high boiling points of ammonia and water and explain how this force originates.
   (c) The shapes of the molecules of hydrogen sulphide and water are similar, yet their boiling points differ. Explain.

   (IEB Paper 2, 2002)

Chapter 6

Solutions and solubility - Grade 11

We are surrounded by different types of solutions in our daily lives. Any solution is made up of a **solute** and a **solvent**. A **solute** is a substance that dissolves in a solvent. In the case of a salt (NaCl) solution, the salt crystals are the solute. A **solvent** is the substance in which the solute dissolves. In the case of the NaCl solution, the solvent would be the water. In most cases, there is always more of the solvent than there is of the solute in a solution.

Definition: Solutes and solvents
A **solute** is a substance that is dissolved in another substance. A solute can be a solid, liquid or gas. A **solvent** is the liquid that dissolves a solid, liquid, or gaseous solute.

6.1 Types of solutions

When a solute is mixed with a solvent, a *mixture* is formed, and this may be either *heterogeneous* or *homogeneous*. If you mix sand and water for example, the sand does not dissolve in the water. This is a **heterogeneous** mixture. When you mix salt and water, the resulting mixture is **homogeneous** because the solute has dissolved in the solvent.

Definition: Solution
In chemistry, a solution is a homogeneous mixture that consists of a solute that has been dissolved in a solvent.

A solution then is a homogeneous mixture of a solute and a solvent. Examples of solutions are:

- A *solid solute* dissolved in a *liquid solvent* e.g. sodium chloride dissolved in water.
- A *gas solute* dissolved in a *liquid solvent* e.g. carbon dioxide dissolved in water (fizzy drinks) or oxygen dissolved in water (aquatic ecosystems).
- A *liquid solute* dissolved in a *liquid solvent* e.g. ethanol in water.
- A *solid solute* in a *solid solvent* e.g. metal alloys.
- A *gas solute* in a *gas solvent* e.g. the homogeneous mixture of gases in the air that we breathe.

While there are many different types of solutions, most of those we will be discussing are *liquids*.

6.2 Forces and solutions

An important question to ask is why some solutes dissolve in certain solvents and not in others. The answer lies in understanding the interaction between the intramolecular and intermolecular forces between the solute and solvent particles.

Activity :: Experiment : Solubility

Aim:

To investigate the solubility of solutes in different solvents.

Apparatus:

Salt, vinegar, iodine, ethanol

Method:

1. Mix half a teaspoon of salt in 100cm^3 of water
2. Mix half a teaspoon of vinegar (acetic acid) in 100cm^3 of water
3. Mix a few grains of iodine in ethanol
4. Mix a few grains of iodine in 100cm^3 of water

Results:

Record your observations in the table below:

Solute	**Polar, non-polar or ionic solute**	**Solvent**	**Polar, non-polar or ionic solvent**	**Does solute dissolve?**
Iodine		Ethanol		
Iodine		Water		
Vinegar		Water		
Salt		Water		

You should have noticed that in some cases, the solute dissolves in the solvent, while in other cases it does not.

Conclusions:

In general, polar and ionic solutes dissolve well in polar solvents, while non-polar solutes dissolve well in non-polar solvents. An easy way to remember this is that 'like dissolves like', in other words, if the solute and the solvent have similar intermolecular forces, there is a high possibility that dissolution will occur. This will be explained in more detail below.

- **Non-polar solutes and non-polar solvents** (e.g. iodine and ether)

  Iodine molecules are non-polar, and the forces between the molecules are weak van der Waals forces. There are also weak van der Waals forces between ether molecules. Because the intermolecular forces in both the solute and the solvent are similar, it is easy for these to be broken in the solute, allowing the solute to move into the spaces between the molecules of the solvent. The solute dissolves in the solvent.

- **Polar solutes and polar solvents** (e.g. salt and water)

  There are strong electrostatic forces between the ions of a salt such as sodium chloride. There are also strong hydrogen bonds between water molecules. Because the strength of the intermolecular forces in the solute and solvent are similar, the solute will dissolve in the solvent.

6.3 Solubility

You may have noticed sometimes that, if you try to dissolve salt (or some other solute) in a small amount of water, it will initially dissolve, but then appears not to be able to dissolve any further when you keep adding more solute to the solvent. This is called the **solubility** of the solution. Solubility refers to the maximum amount of solute that will dissolve in a solvent under certain conditions.

Definition: Solubility
Solubility is the ability of a given substance, the solute, to dissolve in a solvent. If a substance has a high solubility, it means that lots of the solute is able to dissolve in the solvent.

So what factors affect solubility? Below are some of the factors that affect solubility:

- the quantity of solute and solvent in the solution
- the temperature of the solution
- other compounds in the solvent affect solubility because they take up some of the spaces between molecules of the solvent, that could otherwise be taken by the solute itself
- the strength of the forces between particles of the solute, and the strength of forces between particles of the solvent

---

Activity :: Experiment : Factors affecting solubility
Aim:
To determine the effect of temperature on solubility
Method:

1. Measure 100cm^3 of water into a beaker
2. Measure 100 g of salt and place into another beaker
3. Slowly pour the salt into the beaker with the water, stirring it as you add. Keep adding salt until you notice that the salt is not dissolving anymore.
4. Record the amount of salt that has been added to the water and the temperature of the solution.
5. Now increase the temperature of the water by heating it over a bunsen burner.
6. Repeat the steps above so that you obtain the solubility limit of salt at this higher temperature. You will need to start again with new salt and water!
7. Continue to increase the temperature as many times as possible and record your results.

Results:

Record your results in the table below:

Temp (^{0}C)	**Amount of solute that dissolves in 100 cm^3 of water** (g)

As you increase the temperature of the water, are you able to dissolve *more* or *less* salt?

Conclusions:

As the temperature of the solution increases, so does the amount of salt that will dissolve. The solubility of sodium chloride increases as the temperature increases.

?

Exercise: Investigating the solubility of salts

The data table below gives the solubility (measured in grams of salt per 100 g water) of a number of different salts at various temperatures. Look at the data and then answer the questions that follow.

	Solubility (g salt per 100 g H_2O)		
Temp (^{0}C)	**KNO_3**	**K_2SO_4**	**NaCl**
0	13.9	7.4	35.7
10	21.2	9.3	35.8
20	31.6	11.1	36.0
30	45.3	13.0	36.2
40	61.4	14.8	36.5
50	83.5	16.5	36.8
60	106.0	18.2	37.3

1. On the same set of axes, draw line graphs to show how the solubility of the three salts changes with an increase in temperature.
2. Describe what happens to salt solubility as temperature increases. Suggest a reason why this happens.
3. Write an equation to show how each of the following salts ionises in water:
   (a) KNO_3
   (b) K_2SO_4
4. You are given three beakers, each containing the same amount of water. 5 g KNO_3 is added to beaker 1, 5 g K_2SO_4 is added to beaker 2 and 5 g NaCl is added to beaker 3. The beakers are heated over a bunsen burner until the temperature of their solutions is 60^{0}C.
   (a) Which salt solution will have the highest conductivity under these conditions?
   (b) Explain your answer.

?

Exercise: Experiments and solubility

Two grade 10 learners, Siphiwe and Ann, wish to separately investigate the solubility of potassium chloride at room temperature. They follow the list of instructions shown below, using the apparatus that has been given to them:

Method:

1. Determine the mass of an empty, dry evaporating basin using an electronic balance and record the mass.
2. Pour 50 ml water into a 250 ml beaker.

3. Add potassium chloride crystals to the water in the beaker in small portions.
4. Stir the solution until the salt dissolves.
5. Repeat the addition of potassium chloride (steps a and b) until no more salt dissolves and some salt remains undissolved.
6. Record the temperature of the potassium chloride solution.
7. Filter the solution into the evaporating basin.
8. Determine the mass of the evaporating basin containing the solution that has passed through the filter (the filtrate) on the electronic balance and record the mass.
9. Ignite the Bunsen burner.
10. Carefully heat the filtrate in the evaporating basin until the salt is dry.
11. Place the evaporating basin in the desiccator (a large glass container in which there is a dehydrating agent like calcium sulphate that absorbs water) until it reaches room temperature.
12. Determine the mass of the evaporating basin containing the dry cool salt on the electronic balance and record the mass.

On completion of the experiment, their results were as follows:

	Siphiwe's results	Ann's results
Temperature (^{0}C)	15	26
Mass of evaporating basin (g)	65.32	67.55
Mass of evaporating basin + salt solution (g)	125.32	137.55
Mass of evaporating basin + salt (g)	81.32	85.75

1. Calculate the solubility of potassium chloride, using the data recorded by
   (a) Siphiwe
   (b) Ann
   A reference book lists the solubility of potassium chloride as 35.0 g per 100 ml of water at 25^{0}C.
   (c) Give a reason why you think each obtained results different from each other and the value in the reference book.
2. Siphiwe and Ann now expand their investigation and work together. They investigate the solubility of potassium chloride at different temperatures but also the solubility of copper (II) sulphate at these same temperatures. They collect and write up their results as follows:

   In each experiment we used 50 ml of water in the beaker. We found the following masses of substance dissolved in the 50 ml of water. At 0^{0}C, mass of potassium chloride is 14.0 g and copper sulphate is 14.3 g. At 10^{0}C, 15.6 g and 17.4 g respectively. At 20^{0}C, 17.3 g and 20.7 g respectively. At 40^{0}C, potassium chloride mass is 20.2 g and copper sulphate is 28.5 g, at 60^{0}C, 23.1 g and 40.0 g and lastly at 80^{0}C, the masses were 26.4 g and 55.0 g respectively.

   (a) From the record of data provided above, draw up a neat table to record Siphiwe and Ann's results.
   (b) Identify the dependent and independent variables in their investigation.
   (c) Choose an appropriate scale and plot a graph of these results.
   (d) From the graph, determine:
      i. the temperature at which the solubility of copper sulphate is 50 g per 50 ml of water.
      ii. the maximum number of grams of potassium chloride which will dissolve in 100 ml of water at 70^{0}C.

   (IEB Exemplar Paper 2, 2006)

6.4 Summary

- In chemistry, a **solution** is a homogenous mixture of a solute in a solvent.
- A **solute** is a substance that dissolves in a solute. A solute can be a solid, liquid or gas.
- A **solvent** is a substance in which a solute dissolves. A solvent can also be a solid, liquid or gas.
- Examples of solutions include salt solutions, metal alloys, the air we breathe and gases such as oxygen and carbon dioxide dissolved in water.
- Not all solutes will dissolve in all solvents. A general rule is the **like dissolves like**. Solutes and solvents that have similar intermolecular forces are more likely to dissolve.
- Polar and ionic solutes will be more likely to dissolve in polar solvents, while non-polar solutes will be more likely to dissolve in polar solvents.
- **Solubility** is the extent to which a solute is able to dissolve in a solvent under certain conditions.
- Factors that affect solubility are the **quantity of solute and solvent**, **temperature**, the **intermolecular forces** in the solute and solvent and **other substances** that may be in the solvent.

Exercise: Summary Exercise

1. Give one word or term for each of the following descriptions:
   (a) A type of mixture where the solute has completely dissolved in the solvent.
   (b) A measure of how much solute is dissolved in a solution.
   (c) Forces between the molecules in a substance.
2. For each of the following questions, choose the one correct answer from the list provided.
   A Which one of the following will readily dissolve in water?
   i. I_2(s)
   ii. NaI(s)
   iii. CCl_4(l)
   iv. $BaSO_4$(s)
   (IEB Paper 2, 2005)
   b In which of the following pairs of substances will the dissolving process happen most readily?

	Solute	Solvent
A	S_8	H_2O
B	KCl	CCl_4
C	KNO_3	H_2O
D	NH_4Cl	CCl_4

   (IEB Paper 2, 2004)
3. Which one of the following three substances is the most soluble in pure water at room temperature?

   Hydrogen sulphide, ammonia and hydrogen fluoride

4. Briefly explain in terms of intermolecular forces why solid iodine does not dissolve in pure water, yet it dissolves in xylene, an organic liquid at room temperature.
   (IEB Paper 2, 2002)

Chapter 7

Atomic Nuclei - Grade 11

Nuclear physics is the branch of physics which deals with the **nucleus** of the atom. Within this field, some scientists focus their attention on looking at the *particles* inside the nucleus and understanding how they interact, while others classify and interpret the *properties* of nuclei. This detailed knowledge of the nucleus makes it possible for *technological advances* to be made. In this next chapter, we are going to touch on each of these different areas within the field of nuclear physics.

7.1 Nuclear structure and stability

You will remember from an earlier chapter that an atom is made up of different types of particles: protons (positive charge) neutrons (neutral) and electrons (negative charge). The nucleus is the part of the atom that contains the protons and the neutrons, while the electrons are found in energy orbitals around the nucleus. The protons and neutrons together are called **nucleons**. It is the nucleus that makes up most of an atom's *atomic mass*, because an electron has a very small mass when compared with a proton or a neutron.

Within the nucleus, there are different forces which act between the particles. The **strong nuclear force** is the force between two or more nucleons, and this force binds protons and neutrons together inside the nucleus. This force is most powerful when the nucleus is small, and the nucleons are close together. The **electromagnetic force** causes the repulsion between like-charged (positive) protons. In a way then, these forces are trying to produce opposite effects in the nucleus. The strong nuclear force acts to hold all the protons and neutrons close together, while the electromagnetic force acts to push protons further apart. In atoms where the nuclei are small, the strong nuclear force overpowers the electromagnetic force. However, as the nucleus gets bigger (in elements with a higher number of nucleons), the electromagnetic force becomes greater than the strong nuclear force. In these nuclei, it becomes possible for particles and energy to be ejected from the nucleus. These nuclei are called **unstable**. The particles and energy that a nucleus releases are referred to as **radiation**, and the atom is said to be **radioactive**. We are going to look at these concepts in more detail in the next few sections.

7.2 The Discovery of Radiation

Radioactivity was first discovered in 1896 by a French scientist called Henri Becquerel while he was working on phosphorescent materials. He happened to place some uranium crystals on black paper that he had used to cover a piece of film. When he looked more carefully, he noticed that the film had lots of patches on it, and that this did not happen when other elements were placed on the paper. He eventually concluded that some rays must be coming out of the uranium crystals to produce this effect.

His observations were taken further by the Polish scientist Marie Curie and her husband Pierre, who increased our knowledge of radioactive elements. In 1903, Henri, Marie and Pierre were

awarded the Nobel Prize in Physics for their work on radioactive elements. This award made Marie the first woman ever to receive a Nobel Prize. Marie Curie and her husband went on to discover two new elements, which they named **polonium** (Po) after Marie's home country, and **radium** (Ra) after its highly radioactive characteristics. For these dicoveries, Marie was awarded a Nobel Prize in Chemistry in 1911, making her one of very few people to receive two Nobel Prizes.

Interesting Fact

Marie Curie died in 1934 from aplastic anemia, which was almost certainly partly due to her massive exposure to radiation during her lifetime. Most of her work was carried out in a shed without safety measures, and she was known to carry test tubes full of radioactive isotopes in her pockets and to store them in her desk drawers. By the end of her life, not only was she very ill, but her hands had become badly deformed due to their constant exposure to radiation. Unfortunately it was only later in her life that the full dangers of radiation were realised.

7.3 Radioactivity and Types of Radiation

In section 7.1, we discussed that when a nucleus is unstable it can emit particles and energy. This is called **radioactive decay**.

Definition: Radioactive decay
Radioactive decay is the process in which an unstable atomic nucleus loses energy by emitting particles or electromagnetic waves. **Radiation** is the name for the emitted particles or electromagnetic waves.

When a nucleus undergoes radioactive decay, it emits radiation and the nucleus is called radioactive. We are exposed to small amounts of radiation all the time. Even the rocks around us emit radiation! However some elements are far more radioactive than others. *Isotopes* tend to be less stable because they contain a larger number of nucleons than 'non-isotopes' of the same element. These radioactive isotopes are called **radioisotopes**.

Radiation can be emitted in different forms. There are three main types of radiation: alpha, beta and gamma radiation. These are shown in figure 7.1, and are described below.

Figure 7.1: Types of radiation

7.3.1 Alpha (α) particles and alpha decay

An alpha particle is made up of two protons and two neutrons bound together. This type of radiation has a *positive charge*. An alpha particle is sometimes represented using the chemical symbol He^{2+}, because it has the same structure as a Helium atom (two neutrons and two protons) which is missing its two electrons, hence the overall charge of +2. Alpha particles have very low penetration power. Penetration power describes how easily the particles can pass through another material. Because alpha particles have a *low* penetration power, it means that even something as thin as a piece of paper or the outside layer of the human skin, will absorb these particles so that they can't go any further.

Alpha decay occurs because the nucleus has too many protons, and this causes a lot of repulsion between these like charges. To try to reduce this repulsion, the nucleus emits an α particle. This can be seen in the decay of Americium (Am) to Neptunium (Np).

Example:

$$^{241}_{95}\mathrm{Am} \rightarrow ^{237}_{93}\mathrm{Np} + \alpha\mathrm{particle}$$

Let's take a closer look at what has happened during this reaction. Americium (Z = 95; A = 241) undergoes α decay and releases one alpha particle (i.e. 2 protons and 2 neutrons). The atom now has only 93 protons (Z = 93). On the periodic table, the element which has 93 protons (Z = 93) is called Neptunium. Therefore, the Americium atom has become a Neptunium atom. The atomic mass of the neptunium atom is 237 (A = 237) because 4 nucleons (2 protons and 2 neutrons) were emitted from the atom of Americium.

7.3.2 Beta (β) particles and beta decay

In certain types of radioactive nuclei that have too many neutrons, a neutron may be converted into a proton, an electron and another particle (called a *neutrino*). The high energy electrons that are released in this way are called **beta particles**. Beta particles have a higher penetration power than alpha particles and are able to pass through thicker materials such as paper.

The diagram below shows what happens during β decay:

Figure 7.2: β decay in a hydrogen atom

During beta decay, the number of neutrons in the atom decreases by one, and the number of protons increases by one. Since the number of protons before and after the decay is different,

the atom has changed into a different element. In figure 7.2, Hydrogen has become Helium. The beta decay of the Hydrogen-3 atom can be represented as follows:

$$^{3}_{1}\text{H} \rightarrow ^{3}_{2}\text{He} + \beta\text{particle} + \bar{\nu}$$

Interesting Fact

When scientists added up all the energy from the neutrons, protons and electrons involved in β-decays, they noticed that there was always some energy missing. We know that energy is always conserved, which led Wolfgang Pauli in 1930 to come up with the idea that another particle, which was not detected yet, also had to be involved in the decay. He called this particle the neutrino (Italian for "little neutral one"), because he knew it had to be neutral, have little or no mass, and interact only very weakly, making it very hard to find experimentally! The neutrino was finally identified experimentally about 25 years after Pauli first thought of it.

Due to the radioactive processes inside the sun, each 1 cm^2 patch of the earth receives 70 billion (70×10^9) neutrinos each second! Luckily neutrinos only interact very weakly so they do not harm our bodies when billions of them pass through us every second.

7.3.3 Gamma (γ) rays and gamma decay

When particles inside the nucleus collide during radioactive decay, energy is released. This energy can leave the nucleus in the form of waves of electromagnetic energy called gamma rays. Gamma radiation is part of the electromagnetic spectrum, just like visible light. However, unlike visible light, humans cannot see gamma rays because they are at a higher frequency and a higher energy. Gamma radiation has no mass or charge. This type of radiation is able to penetrate most common substances, including metals. The only substances that can absorb this radiation are thick lead and concrete.

Gamma decay occurs if the nucleus is at too high an energy level. Since gamma rays are part of the electromagnetic spectrum, they can be thought of as waves *or* particles. Therefore in gamma decay, we can think of a ray or a particle (called a photon) being released. The atomic number and atomic mass remain unchanged.

Table 7.1 summarises and compares the three types of radioactive decay that have been discussed.

Table 7.1: A comparison of alpha, beta and gamma decay

Type of decay	**Particle/ray released**	**Change in element**	**Penetration power**
Alpha (α)	α particle (2 protons and 2 neutrons)	Yes	Low
Beta (β)	β particle (electron)	Yes	Medium
Gamma (γ)	γ ray (electromagnetic energy)	No	High

Worked Example 23: Radioactive decay

Question: The isotope $^{241}_{95}$Pb undergoes radioactive decay and loses two alpha particles.

Figure 7.3: γ decay in a helium atom

1. Write the chemical formula of the element that is produced as a result of the decay.
2. Write an equation for this decay process.

Answer
Step 1 : Work out the number of protons and/or neutrons that the radioisotope loses during radioactive decay
One α particle consists of two protons and two neutrons. Since two α particles are released, the total number of protons lost is four and the total number of neutrons lost is also four.

Step 2 : Calculate the atomic number (Z) and atomic mass number (A) of the element that is formed.

$$Z = 95 - 4 = 91$$

$$A = 241 - 4 = 237$$

Step 3 : Refer to the periodic table to see which element has the atomic number that you have calculated.
The element that has Z = 91 is Protactinium (Pa).

Step 4 : Write the symbol for the element that has formed as a result of radioactive decay.

$$^{237}_{91}\text{Pa}$$

Step 5 : Write an equation for the decay process.

$$^{241}_{95}Pb \rightarrow ^{237}_{91}Pa + 2 \text{ protons} + 2 \text{ neutrons}$$

Activity :: Discussion : Radiation
In groups of 3-4, discuss the following questions:

- Which of the three types of radiation is most dangerous to living creatures (including humans!)

- What can happen to people if they are exposed to high levels of radiation?
- What can be done to protect yourself from radiation (Hint: Think of what the radiologist does when you go for an X-ray)?

Exercise: Radiation and radioactive elements

1. There are two main forces inside an atomic nucleus:
   (a) Name these two forces.
   (b) Explain why atoms that contain a greater number of nucleons are more likely to be radioactive.
2. The isotope $^{241}_{95}$Pb undergoes radioactive decay and loses three alpha particles.
   (a) Write the chemical formula of the element that is produced as a result of the decay.
   (b) How many nucleons does this element contain?
3. Complete the following equation:
   $$^{210}_{82}Am \rightarrow \text{(alpha decay)}$$
4. Radium-228 decays by emitting a beta particle. Write an equation for this decay process.
5. Describe how gamma decay differs from alpha and beta decay.

7.4 Sources of radiation

The sources of radiation can be either **natural** or **man-made**.

7.4.1 Natural background radiation

- *Cosmic radiation*

  The Earth, and all living things on it, are constantly bombarded by radiation from space. Charged particles from the sun and stars interact with the Earth's atmosphere and magnetic field to produce a shower of radiation, mostly beta and gamma radiation. The amount of cosmic radiation varies in different parts of the world because of differences in elevation and also the effects of the Earth's magnetic field.

- *Terrestrial Radiation*

  Radioactive material is found throughout nature. It occurs naturally in the soil, water, and vegetation. The major isotopes that are of concern are uranium and the decay products of uranium, such as thorium, radium, and radon. Low levels of uranium, thorium, and their decay products are found everywhere. Some of these materials are ingested (taken in) with food and water, while others are breathed in. The dose of radiation from terrestrial sources varies in different parts of the world.

Interesting Fact

Cosmic and terrestrial radiation are not the only natural sources. All people have radioactive potassium-40, carbon-14, lead-210 and other isotopes inside their bodies from birth.

7.4.2 Man-made sources of radiation

Although all living things are exposed to natural background radiation, there are other sources of radiation. Some of these will affect most members of the public, while others will only affect those people who are exposed to radiation through their work.

- *Members of the Public*

  Man-made radiation sources that affect members of the public include televisions, tobacco (polonium-210), combustible fuels, smoke detectors (americium), luminous watches (tritium) and building materials. By far, the most significant source of man-made radiation exposure to the public is from medical procedures, such as diagnostic x-rays, nuclear medicine, and radiation therapy. Some of the major isotopes involved are I-131, Tc-99m, Co-60, Ir-192, and Cs-137. The production of nuclear fuel using uranium is also a source of radiation for the public, as is fallout from nuclear weapons testing or use.

- *Individuals who are exposed through their work*

  Any people who work in the following environments are exposed to radiation at some time: radiology (X-ray) departments, nuclear power plants, nuclear medicine departments and radiation oncology (the study of cancer) departments. Some of the isotopes that are of concern are cobalt-60, cesium-137, americium-241, and others.

Interesting Fact

Radiation therapy (or radiotherapy) uses ionising radiation as part of cancer treatment to control malignant cells. In cancer, a malignant cell is one that divides very rapidly to produce many more cells. These groups of dividing cells can form a growth or **tumour**. The malignant cells in the tumour can take nutrition away from other healthy body cells, causing them to die, or can increase the pressure in parts of the body because of the space that they take up. Radiation therapy uses radiation to try to target these malignant cells and kill them. However, the radiation can also damage other, healthy cells in the body. To stop this from happening, shaped radiation beams are aimed from several angles to intersect at the tumour, so that the radiation dose here is much higher than in the surrounding, healthy tissue. But even doing this doesn't protect all the healthy cells, and that is why people have side-effects to this treatment.

Note that radiation therapy is different from chemotherapy, which uses *chemicals*, rather than radiation, to destroy malignant cells. Generally, the side effects of chemotherapy are greater because the treatment is not as localised as it is with radiation therapy. The chemicals travel throughout the body, affecting many healthy cells.

7.5 The 'half-life' of an element

Definition: Half-life
The half-life of an element is the time it takes for half the atoms of a radioisotope to decay into other atoms.

Table 7.2 gives some examples of the half-lives of different elements.

Table 7.2: Table showing the half-life of a number of elements

Radioisotope	Chemical symbol	Half-life
Polonium-212	Po-212	0.16 seconds
Sodium-24	Na-24	15 hours
Strontium-90	Sr-90	28 days
Cobalt-60	Co-60	5.3 years
Caesium-137	Cs-137	30 years
Carbon-14	C-14	5 760 years
Calcium	Ca	100 000 years
Beryllium	Be	2 700 000 years
Uranium-235	U-235	7.1 billion years

So, in the case of Sr-90, it will take 28 days for half of the atoms to decay into other atoms. It will take another 28 days for half of the remaining atoms to decay. Let's assume that we have a sample of strontium that weighs 8g. After the first 28 days there will be:

$$1/2 \text{ x } 8 = 4 \text{ g Sr-90 left}$$

After 56 days, there will be:

$$1/2 \text{ x } 4 \text{ g} = 2 \text{ g Sr-90 left}$$

After 84 days, there will be:

$$1/2 \text{ x } 2 \text{ g} = 1 \text{ g Sr-90 left}$$

If we convert these amounts to a *fraction* of the original sample, then after 28 days 1/2 of the sample remains undecayed. After 56 days 1/4 is undecayed and after 84 days, 1/8 and so on.

Activity :: Group work : Understanding half-life

Work in groups of 4-5

You will need:

16 sheets of A4 paper per group, scissors, 2 boxes per group, a marking pen and timer/stopwatch.

What to do:

- Your group should have two boxes. Label one 'decayed' and the other 'radioactive'.
- Take the A4 pages and cut each into 4 pieces of the same size. You should now have 64 pieces of paper. Stack these neatly and place them in the 'radioactive' box. The paper is going to represent some radioactive material.
- Set the timer for one minute. After one minute, remove half the sheets of paper from the radioactive box and put them in the 'decayed' box.
- Set the timer for another minute and repeat the previous step, again removing half the pieces of paper that are left in the radioactive box and putting them in the decayed box.
- Repeat this process until 8 minutes have passed. You may need to start cutting your pieces of paper into even smaller pieces as you progress.

Questions:

1. How many pages were left in the radioactive box after...
   (a) 1 minute
   (b) 3 minutes

   (c) 5 minutes
2. What percentage (%) of the pages were left in the radioactive box after...
   (a) 2 minutes
   (b) 4 minutes
3. After how many minutes is there 1/128 of radioactive material remaining?
4. What is the half-life of the 'radioactive' material in this exercise?

Worked Example 24: Half-life 1

Question: A 100 g sample of Cs-137 is allowed to decay. Calculate the mass of Cs-137 that will be left after 90 years

Answer
Step 1 : You need to know the half-life of Cs-137
The half-life of Cs-137 is 30 years.

Step 2 : Determine how many times the quantity of sample will be halved in 90 years.
If the half-life of Cs-137 is 30 years, and the sample is left to decay for 90 years, then the number of times the quantity of sample will be halved is 90/30 = 3.

Step 3 : Calculate the quantity that will be left by halving the mass of Cs-137 three times
1. After 30 years, the mass left is 100 g $\times$ 1/2 = 50 g
2. After 60 years, the mass left is 50 g $\times$ 1/2 = 25 g
3. After 90 years, the mass left is 25 g $\times$ 1/2 = 12.5 g

Note that a quicker way to do this calculation is as follows:
Mass left after 90 years = $(1/2)^3$ $\times$ 100 g = 12.5 g (The exponent is the number of times the quantity is halved)

Worked Example 25: Half-life 2

Question: An 80 g sample of Po-212 decays until only 10 g is left. How long did it take for this decay to take place?

Answer
Step 1 : Calculate the fraction of the original sample that is left after decay
Fraction remaining = 10 g/80 g = 1/8

Step 2 : Calculate how many half-life periods of decay (x) must have taken place for 1/8 of the original sample to be left

$$(\frac{1}{2})^x = \frac{1}{8}$$

Therefore, x = 3

Step 3 : Use the half-life of Po-212 to calculate how long the sample was left to decay
The half-life of Po-212 is 0.16 seconds. Therefore if there were three periods of decay, then the total time is 0.16 × 3. The time that the sample was left to decay is 0.48 seconds.

Exercise: Looking at half life

1. Imagine that you have 100 g of Na-24.
   (a) What is the half life of Na-24?
   (b) How much of this isotope will be left after 45 hours?
   (c) What percentage of the original sample will be left after 60 hours?
2. A sample of Sr-90 is allowed to decay. After 84 days, 10 g of the sample remains.
   (a) What is the half life of Sr-90?
   (b) How much Sr-90 was in the original sample?
   (c) How much Sr-90 will be left after 112 days?

7.6 The Dangers of Radiation

Natural radiation comes from a variety of sources such as the rocks, sun and from space. However, when we are exposed to large amounts of radiation, this can cause damage to cells. γ radiation is particularly dangerous because it is able to penetrate the body, unlike α and β particles whose penetration power is less. Some of the dangers of radiation are listed below:

- **Damage to cells**

  Radiation is able to penetrate the body, and also to penetrate the membranes of the cells within our bodies, causing massive damage. *Radiation poisoning* occurs when a person is exposed to large amounts of this type of radiation. Radiation poisoning damages tissues within the body, causing symptoms such as diarrhoea, vomiting, loss of hair and convulsions.

- **Genetic abnormalities**

  When radiation penetrates cell membranes, it can damage chromosomes within the nucleus of the cell. The chromosomes contain all the genetic information for that person. If the chromosomes are changed, this may lead to genetic abnormalities in any children that are born to the person who has been exposed to radiation. Long after the nuclear disaster of Chernobyl in Russia in 1986, babies were born with defects such as missing limbs and abnormal growths.

- **Cancer**

  Small amounts of radiation can cause cancers such as leukemia (cancer of the blood)

7.7 The Uses of Radiation

However, despite the many dangers of radiation, it does have many powerful uses, some of which are listed below:

- **Medical Field**

  Radioactive *chemical tracers* emitting γ rays can give information about a person's internal anatomy and the functioning of specific organs. The radioactive material may be injected into the patient, from where it will target specific areas such as bones or tumours. As the material decays and releases radiation, this can be seen using a special type of camera or other instrument. The radioactive material that is used for this purpose must have a short half-life so that the radiation can be detected quickly and also so that the material is quickly removed from the patient's body. Using radioactive materials for this purpose can mean that a tumour or cancer may be diagnosed long before these would have been detected using other methods such as X-rays.

  Radiation may also be used to sterilise medical equipment.

Activity :: Research Project : The medical uses of radioisotopes

Carry out your own research to find out more about the radioisotopes that are used to diagnose diseases in the following parts of the body:

- thyroid gland
- kidneys
- brain

In each case, try to find out...

1. which radioisotope is used
2. what the sources of this radioisotope are
3. how the radioisotope enters the patient's body and how it is monitored

---

- **Biochemistry and Genetics**

  Radioisotopes may be used as tracers to label molecules so that chemical processes such as DNA replication or amino acid transport can be traced.

- **Food preservation**

  Irradiation of food can stop vegetables or plants from sprouting after they have been harvested. It also kills bacteria and parasites, and controls the ripening of fruits.

- **Environment**

  Radioisotopes can be used to trace and analyse pollutants.

- **Archaeology and Carbon dating**

  Natural radioisotopes such as C-14 can be used to determine the age of organic remains. All living organisms (e.g. trees, humans) contain carbon. Carbon is taken in by plants and trees through the process of photosynthesis in the form of carbon dioxide and is then converted into organic molecules. When animals feed on plants, they also obtain carbon through these organic compounds. Some of the carbon in carbon dioxide is the radioactive C-14, while the rest is a non-radioactive form of carbon. When an organism dies, no more carbon is taken in and there is a limited amount of C-14 in the body. From this point onwards, C-14 begins its radioactive decay. When scientists uncover remains, they are able to estimate the age of the remains by seeing how much C-14 is left in the body relative to the amount of non-radioactive carbon. The less C-14 there is, the older the remains because radioactive decay must have been taking place for a long time. Because scientists know the exact rate of decay of C-14, they can calculate a very accurate estimate of the age of the remains. Carbon dating has been a very important tool in building up accurate historical records.

Activity :: Case Study : Using radiocarbon dating

Radiocarbon dating has played an important role in uncovering many aspects of South Africa's history. Read the following extract from an article that appeared in Afrol news on 10th February 2007 and then answer the questions that follow.

> The world famous rock art in South Africa's uKhahlamba-Drakensberg, a World Heritage Site, is three times older than previously thought, archaeologists conclude in a new study. The more than 40,000 paintings were made by the San people some 3000 years ago, a new analysis had shown.
> Previous work on the age of the rock art in uKhahlamba-Drakensberg concluded it is less than 1,000 years old. But the new study - headed by a South African archaeologist leading a team from the University of Newcastle upon Tyne (UK) and Australian National University in Canberra - estimates the panels were created up to 3,000 years ago. They used the latest radio-carbon dating technology.
> The findings, published in the current edition of the academic journal 'South African Humanities', have "major implications for our understanding of how the rock artists lived and the social changes that were taking place over the last three millennia," according to a press release from the British university.

Questions:

1. What is the half-life of carbon-14?
2. In the news article, what role did radiocarbon dating play in increasing our knowledge of South Africa's history?
3. Radiocarbon dating can also be used to analyse the remains of once-living organisms. Imagine that a set of bones are found between layers of sediment and rock in a remote area. A group of archaeologists carries out a series of tests to try to estimate the age of the bones. They calculate that the bones are approximately 23 040 years old.
   What percentage of the original carbon-14 must have been left in the bones for them to arrive at this estimate?

7.8 Nuclear Fission

Nuclear fission is a process where the nucleus of an atom is split into two or more smaller nuclei, known as *fission products*. The fission of heavy elements is an **exothermic reaction** and huge amounts of energy are released in the process. This energy can be used to produce *nuclear power* or to make *nuclear weapons*, both of which we will discuss a little later.

Definition: Nuclear fission
The splitting of an atomic nucleus

Below is a diagram showing the nuclear fission of Uranium-235. An atom of Uranium-235 is bombarded with a neutron to initiate the fission process. This neutron is absorbed by Uranium-235, to become Uranium-236. Uranium-236 is highly unstable and breaks down into a number of lighter elements, releasing energy in the process. Free neutrons are also produced during this process, and these are then available to bombard other fissionable elements. This process is known as a **fission chain reaction**, and occurs when one nuclear reaction starts off another, which then also starts off another one so that there is a rapid increase in the number of nuclear reactions that are taking place.

7.8.1 The Atomic bomb - an abuse of nuclear fission

A nuclear chain reaction can happen very quickly, releasing vast amounts of energy in the process. In 1939, it was discovered that Uranium could undergo nuclear fission. In fact, it was uranium that was used in the first atomic bomb. The bomb contained huge amounts of Uranium-235, enough to start a runaway nuclear fission chain reaction. Because the process was uncontrolled, the energy from the fission reactions was released in a matter of *seconds*, resulting in the massive explosion of that first bomb. Since then, more atomic bombs have been dropped, causing massive destruction and loss of life.

Activity :: Discussion : Nuclear weapons testing - an ongoing issue

Read the article below which has been adapted from one that appeared in 'The Globe' in Washington on 10th October 2006, and then answer the questions that follow.

> US officials and arms control specialists warned yesterday that North Korea's test of a small nuclear device could start an arms race in the region and threaten the landmark global treaty designed nearly four decades ago to halt the spread of nuclear weapons. US officials expressed concern that North Korea's neighbors, including Japan, Taiwan, and South Korea, could eventually decide to develop weapons of their own. They also fear that North Korea's moves could embolden Iran, and that this in turn could encourage Saudi Arabia or other neighbours in the volatile Middle East to one day seek nuclear deterrents, analysts say.
>
> North Korea is the first country to conduct a nuclear test after pulling out of the Nuclear Nonproliferation Treaty. The treaty, which was created in 1968, now includes 185 nations (nearly every country in the world). Under the treaty, the five declared nuclear powers at the time (United States, the Soviet Union, France, China, and Great Britain) agreed to reduce their supplies of nuclear weapons. The treaty has also helped to limit the number of new nuclear weapons nations.
>
> But there have also been serious setbacks. India and Pakistan, which never signed the treaty, became new nuclear powers, shocking the world with test explosions in 1998. The current issue of nuclear weapons testing in North Korea, is another such setback and a blow to the treaty.

Group discussion questions:

1. Discuss what is meant by an 'arms race' and a 'treaty'.

2. Do you think it is important to have such treaties in place to control the testing and use of nuclear weapons? Explain your answer.
3. Discuss some of the reasons why countries might not agree to be part of a nuclear weapons treaty.
4. How would you feel if South Africa decided to develop its own nuclear weapons?

7.8.2 Nuclear power - harnessing energy

However, nuclear fission can also be carried out in a controlled way in a *nuclear reactor*. A nuclear reactor is a piece of equpiment where nuclear chain reactions can be started in a controlled and sustained way. This is different from a nuclear *explosion* where the chain reaction occurs in seconds. The most important use of nuclear reactors at the moment is to produce **electrical power**, and most of these nuclear reactors use nuclear fission. A **nuclear fuel** is a chemical isotope that can keep a fission chain reaction going. The most common isotopes that are used are Uranium-235 and Plutonium-239. The amount of free energy that is in nuclear fuels is far greater than the energy in a similar amount of other fuels such as gasoline. In many countries, nuclear power is seen as a relatively environmentally friendly alternative to fossil fuels, which release large amounts of greenhouse gases, and are also non-renewable resources. However, one of the concerns around the use of nuclear power, is the production of *nuclear waste* which contains radioactive chemical elements.

Activity :: Debate : Nuclear Power

The use of nuclear power as a source of energy has been a subject of much debate. There are many advantages of nuclear power over other energy sources. These include the large amount of energy that can be produced at a small plant, little atmospheric pollution and the small quantity of waste. However there are also disadvantages. These include the expense of maintaining nuclear power stations, the huge impact that an accident could have as well as the disposal of dangerous nuclear waste.

Use these ideas as a starting point for a class debate.

Nuclear power - An energy alternative or environmental hazard?

Your teacher will divide the class into teams. Some of the teams will be 'pro' nuclear power while the others will be 'anti' nuclear power.

7.9 Nuclear Fusion

Nuclear fusion is the joining together of the nuclei of two atoms to form a heavier nucleus. If the atoms involved are small, this process is accompanied by the release of energy. It is the nuclear fusion of elements that causes stars to shine and hydrogen bombs to explode. As with nuclear *fission* then, there are both positive and negative uses of nuclear fusion.

Definition: Nuclear fusion
The joining together of the nuclei of two atoms.

You will remember that nuclei naturally repel one another because of the electrostatic force between their positively charged protons. So, in order to bring two nuclei together, a lot of energy must be supplied if fusion is to take place. If two nuclei can be brought close enough together however, the electrostatic force is overwhelmed by the more powerful strong nuclear force which only operates over short distances. If this happens, nuclear fusion can take place. Inside the cores of stars, the temperature is high enough for hydrogen fusion to take place but scientists have so far been unsuccessful in making it work in the laboratory. One of the huge advantages of nuclear fusion, if it could be made to happen, is that it is a relatively environmentally friendly source of energy. The helium that is produced is not radioactive or poisonous and does not carry the dangers of nuclear fission.

7.10 Nucleosynthesis

An astronomer named Edwin Hubble discovered in the 1920's that the universe is expanding. He measured that far-away galaxies are moving away from the earth at great speed, and the further away they are, the faster they are moving.

Extension: What are galaxies?
Galaxies are huge clusters of stars and matter in the universe. The earth is part of the Milky Way galaxy which is shaped like a very large spiral. Astronomers can measure the light coming from distant galaxies using telescopes. Edwin Hubble was also able to measure the velocities of galaxies.

These observations led people to see that the universe is expanding. It also led to the *Big Bang* hypothesis. The 'Big Bang' hypothesis is an idea about how the universe may have started. According to this theory, the universe started off at the beginning of time as a point which then exploded and expanded into the universe we live in today. This happened between 10 and 14 billion years ago.

Just after the Big Bang, when the universe was only 10^{-43}s old, it was very hot and was made up of quarks and leptons (an example of a lepton is the electron). As the universe expanded, ($\sim 10^{-2}$s) and cooled, the quarks started binding together to form protons and neutrons (together called *nucleons*).

7.10.1 Age of Nucleosynthesis (225 s - 10^3 s)

About 225 s after the Big Bang, the protons and neutrons started binding together to form simple *nuclei*. The process of forming nuclei is called *nucleosynthesis*. When a proton and a neutron bind together, they form the *deuteron*. The deuteron is like a hydrogen nucleus (which is just a proton) with a neutron added to it so it can be written as $^2\mathrm{H}$. Using protons and neutrons as building blocks, more nuclei can be formed as shown below. For example, the Helium-4 nucleus (also called an *alpha particle*) can be formed in the following ways:

$$\begin{aligned} {}^2\mathrm{H} + \mathrm{n} &\rightarrow {}^3\mathrm{H} \\ \text{deuteron} + \text{neutron} &\rightarrow \text{triton} \end{aligned}$$

then:

$$\begin{aligned} {}^3\mathrm{H} + \mathrm{p} &\rightarrow {}^4\mathrm{He} \\ \text{triton} + \text{proton} &\rightarrow \text{Helium4 (alpha particle)} \end{aligned}$$

or

$$\begin{aligned} {}^{2}\mathrm{H} + \mathrm{p} &\rightarrow {}^{3}\mathrm{He} \\ \text{deuteron} + \text{proton} &\rightarrow \text{Helium3} \end{aligned}$$

then:

$$\begin{aligned} {}^{3}\mathrm{He} + \mathrm{n} &\rightarrow {}^{4}\mathrm{He} \\ \text{Helium3} + \text{neutron} &\rightarrow \text{Helium4 (alpha particle)} \end{aligned}$$

Some ${}^{7}\mathrm{Li}$ nuclei could also have been formed by the fusion of ${}^{4}\mathrm{He}$ and ${}^{3}\mathrm{H}$.

7.10.2 Age of Ions (10^3 s - 10^{13} s)

However, at this time the universe was still very hot and the electrons still had too much energy to become bound to the alpha particles to form helium *atoms*. Also, the nuclei with mass numbers greater than 4 (i.e. greater than ${}^{4}\mathrm{He}$) are very short-lived and would have decayed almost immediately after being formed. Therefore, the universe moved through a stage called the Age of Ions when it consisted of free positively charged H^{+} ions and ${}^{4}\mathrm{He}$ ions, and negatively charged electrons not yet bound into atoms.

7.10.3 Age of Atoms (10^{13} s - 10^{15} s)

As the universe expanded further, it cooled down until the electrons were able to bind to the hydrogen and helium nuclei to form hydrogen and helium atoms. Earlier, during the Age of Ions, both the hydrogen and helium ions were positively charged which meant that they repelled each other (electrostatically). During the Age of Atoms, the hydrogen and helium along with the electrons, were in the form of atoms which are electrically neutral and so they no longer repelled each other and instead pulled together under gravity to form clouds of gas, which evetually formed stars.

7.10.4 Age of Stars and Galaxies (the universe today)

Inside the core of stars, the densities and temperatures are high enough for fusion reactions to occur. Most of the heavier nuclei that exist today were formed inside stars from thermonuclear reactions! (It's interesting to think that the atoms that we are made of were actually manufactured inside stars!). Since stars are mostly composed of hydrogen, the first stage of thermonuclear reactions inside stars involves hydrogen and is called **hydrogen burning**. The process has three steps and results in four hydrogen atoms being formed into a helium atom with (among other things) two photons (light!) being released.

The next stage is **helium burning** which results in the formation of carbon. All these reactions release a large amount of energy and heat the star which causes heavier and heavier nuclei to fuse into nuclei with higher and higher atomic numbers. The process stops with the formation of ${}^{56}\mathrm{Fe}$, which is the most strongly bound nucleus. To make heavier nuclei, even higher energies are needed than is possible inside normal stars. These nuclei are most likely formed when huge amounts of energy are released, for example when stars explode (an exploding star is called a **supernova**). This is also how all the nuclei formed inside stars get "recycled" in the universe to become part of new stars and planets.

7.11 Summary

- **Nuclear physics** is the branch of physics that deals with the nucleus of an atom.

- There are two forces between the particles of the nucleus. The **strong nuclear force** is an attractive force between the neutrons and the **electromagnetic force** is the repulsive force between like-charged protons.
- In atoms with large nuclei, the electromagnetic force becomes greater than the strong nuclear force and particles or energy may be released from the nucleus.
- **Radioactive decay** occurs when an unstable atomic nucleus loses energy by emitting particles or electromagnetic waves.
- The particles and energy released are called **radiation** and the atom is said to be **radioactive**.
- Radioactive isotopes are called **radioisotopes**.
- Radioactivity was first discovered by Marie Curie and her husband Pierre.
- There are three types of radiation from radioactive decay: **alpha (α)**, **beta (β)** and **gamma (γ)** radiation.
- During **alpha decay**, an alpha particle is released. An alpha particle consists of two protons and two neutrons bound together. Alpha radiation has low penetration power.
- During **beta decay**, a beta particle is released. During beta decay, a neutron is converted to a proton, an electron and a neutrino. A beta particle is the electron that is released. Beta radiation has greater penetration power than alpha radiation.
- During **gamma decay**, electromagnetic energy is released as gamma rays. Gamma radiation has the highest penetration power of the three radiation types.
- There are many **sources of radiation**. Some of natural and others are man-made.
- Natural sources of radiation include cosmic and terrestrial radiation.
- Man-made sources of radiation include televisions, smoke detectors, X-rays and radiation therapy.
- The **half-life** of an element is the time it takes for half the atoms of a radioisotope to decay into other atoms.
- Radiation can be very damaging. Some of the **negative impacts of radiation** exposure include damage to cells, genetic abnormalities and cancer.
- However, radiation can also have many **positive uses**. These include use in the medical field (e.g. chemical tracers), biochemistry and genetics, use in food preservation, the environment and in archaeology.
- **Nuclear fission** is the splitting of an atomic nucleus into smaller fission products. Nuclear fission produces large amounts of energy, which can be used to produce nuclear power, and to make nuclear weapons.
- **Nuclear fusion** is the joining together of the nuclei of two atoms to form a heavier nucleus. In stars, fusion reactions involve the joining of hydrogen atoms to form helium atoms.
- **Nucleosynthesis** is the process of forming nuclei. This was very important in helping to form the universe as we know it.

?

Exercise: Summary exercise

1. Explain each of the following terms:
   (a) electromagnetic force
   (b) radioactive decay
   (c) radiocarbon dating

2. For each of the following questions, choose the **one correct answer**:
   (a) The part of the atom that undergoes radioactive decay is the...
      i. neutrons
      ii. nucleus
      iii. electrons
      iv. entire atom
   (b) The radio-isotope Po-212 undergoes alpha decay. Which of the following statements is **true**?
      i. The number of protons in the element remains unchanged.
      ii. The number of nucleons after decay is 212.
      iii. The number of protons in the element after decay is 82.
      iv. The end product after decay is Po-208.
3. 20 g of sodium-24 undergoes radoactive decay. Calculate the percentage of the original sample that remains after 60 hours.
4. Nuclear physics can be controversial. Many people argue that studying the nucleus has led to devastation and huge loss of life. Others would argue that the benefits of nuclear physics far outweigh the negative things that have come from it.
   (a) Outline some of the ways in which nuclear physics has been used in negative ways.
   (b) Outline some of the benefits that have come from nuclear physics.

Chapter 8

Thermal Properties and Ideal Gases - Grade 11

We are surrounded by gases in our atmosphere which support and protect life on this planet. In this chapter, we are going to try to understand more about gases, and learn how to predict how they will behave under different conditions. The kinetic theory of matter was discussed in chapter 2. This theory is very important in understanding how gases behave.

8.1 A review of the kinetic theory of matter

The main assumptions of the kinetic theory of matter are as follows:

- Matter is made up of **particles** (e.g. atoms or molecules)
- These particles are constantly moving because they have kinetic energy. The space in which the particles move is the **volume** of the gas.
- There are **spaces** between the particles
- There are **attractive forces** between particles and these become stronger as the particles move closer together.
- All particles have **energy**. The **temperature** of a substance is a measure of the **average kinetic energy** of the particles.
- A change in **phase** may occur when the energy of the particles is changed.

The kinetic theory applies to all matter, including gases. In a gas, the particles are far apart and have a high kinetic energy. They move around freely, colliding with each other or with the sides of the container if the gas is enclosed. The **pressure** of a gas is a measure of the frequency of collisions of the gas particles with each other and with the sides of the container that they are in. If the gas is heated, the average kinetic energy of the gas particles will increase and if the temperature is decreased, so does their energy. If the energy of the particles decreases significantly, the gas liquifies. An **ideal gas** is one that obeys all the assumptions of the kinetic theory of matter. A **real gas** behaves like an ideal gas, except at high pressures and low temperatures. This will be discussed in more detail later in this chapter.

Definition: Ideal gas
An ideal gas or perfect gas is a hypothetical gas that obeys all the assumptions of the kinetic theory of matter. In other words, an ideal gas would have identical particles of zero volume, with no intermolecular forces between them. The atoms or molecules in an ideal gas would also undergo elastic collisions with the walls of their container.

Definition: Real gas
Real gases behave more or less like ideal gases except under certain conditions e.g. high pressures and low temperatures.

There are a number of laws that describe how gases behave. It will be easy to make sense of these laws if you understand the kinetic theory of gases that was discussed above.

8.2 Boyle's Law: Pressure and volume of an enclosed gas

Activity :: Demonstration : Boyle's Law

If you have ever tried to force in the plunger of a syringe or a bicycle pump while sealing the opening with a finger, you will have seen Boyle's Law in action! This will now be demonstrated using a 10 ml syringe.

Aim:

To demonstrate Boyle's law.

Apparatus:

You will only need a syringe for this demonstration.

Method:

1. Hold the syringe in one hand, and with the other pull the plunger out towards you so that the syringe is now full of air.
2. Seal the opening of the syringe with your finger so that no air can escape the syringe.
3. Slowly push the plunger in, and notice whether it becomes *more* or *less* difficult to push the plunger in.

Results:

What did you notice when you pushed the plunger in? What happens to the **volume** of air inside the syringe? Did it become *more* or *less* difficult to push the plunger in as the volume of the air in the syringe decreased? In other words, did you have to apply more or less **pressure** to the plunger as the volume of air in the syringe decreased?

As the volume of air in the syringe decreases, you have to apply more pressure to the plunger to keep forcing it down. The pressure of the gas inside the syringe pushing back on the plunger is greater. Another way of saying this is that as the volume of the gas in the syringe *decreases*, the pressure of that gas *increases*.

Conclusion:

If the volume of the gas decreases, the pressure of the gas increases. If the volume of the gas increases, the pressure decreases. These results support Boyle's law.

In the previous demonstration, the volume of the gas decreased when the pressure increased, and the volume increased when the pressure decreased. This is called an **inverse relationship**. The inverse relationship between pressure and volume is shown in figure 8.1.

Figure 8.1: Graph showing the inverse relationship between pressure and volume

Can you use the kinetic theory of gases to explain this inverse relationship between the pressure and volume of a gas? Let's think about it. If you decrease the volume of a gas, this means that the same number of gas particles are now going to come into contact with each other and with the sides of the container much more often. You may remember from earlier that we said that *pressure* is a measure of the *frequency of collisions* of gas particles with each other and with the sides of the container they are in. So, if the volume decreases, the pressure will naturally increase. The opposite is true if the volume of the gas is increased. Now, the gas particles collide less frequently and the pressure will decrease.

It was an Englishman named Robert Boyle who was able to take very accurate measurements of gas pressures and volumes using excellent vacuum pumps. He discovered the startlingly simple fact that the pressure and volume of a gas are not just vaguely inversely related, but are *exactly* **inversely proportional**. This can be seen when a graph of pressure against the inverse of volume is plotted. When the values are plotted, the graph is a straight line. This relationship is shown in figure 8.2.

Figure 8.2: The graph of pressure plotted against the inverse of volume, produces a straight line. This shows that pressure and volume are exactly inversely proportional.

Definition: Boyle's Law
The pressure of a fixed quantity of gas is inversely proportional to the volume it occupies so long as the temperature remains constant.

Important: Proportionality

During this chapter, the terms **directly proportional** and **inversely proportional** will be used a lot, and it is important that you understand their meaning. Two quantities are said to be **proportional** if they vary in such a way that one of the quantities is a constant multiple of the other, or if they have a constant ratio. We will look at two examples to show the difference between *directly proportional* and *inversely proportional*.

1. *Directly proportional*
   A car travels at a constant speed of 120 km/h. The time and the distance covered are shown in the table below.

Time (mins)	**Distance** (km)
10	20
20	40
30	60
40	80

   What you will notice is that the two quantities shown are constant multiples of each other. If you divide each distance value by the time the car has been driving, you will always get 2. This shows that the values are proportional to each other. They are **directly proportional** because both values are increasing. In other words, as the driving time increases, so does the distance covered. The same is true if the values decrease. The shorter the driving time, the smaller the distance covered. This relationship can be described mathematically as:

$$y = kx$$

   where y is distance, x is time and k is the *proportionality constant*, which in this case is 2. Note that this is the equation for a straight line graph! The symbol $\propto$ is also used to show a directly proportional relationship.

2. *Inversely proportional*
   Two variables are inversely proportional if one of the variables is directly proportional to the multiplicative inverse of the other. In other words,

$$y \propto \frac{1}{x}$$

   or

$$y = \frac{k}{x}$$

   This means that as one value gets bigger, the other value will get smaller. For example, the time taken for a journey is inversely proportional to the speed of travel. Look at the table below to check this for yourself. For this example, assume that the distance of the journey is 100 km.

Speed (km/h)	**Time** (mins)
100	60
80	75
60	100
40	150

   According to our definition, the two variables are inversely proportional is one variable is *directly* proportional to the *inverse* of the other. In other words, if we divide one of the variables by the inverse of the other, we should always get the same number. For example,

$$\frac{100}{1/60} = 6000$$

   If you repeat this using the other values, you will find that the answer is always 6000. The variables are inversely proportional to each other.

We know now that the pressure of a gas is *inversely proportional* to the volume of the gas, provided the temperature stays the same. We can write this relationship symbolically as

$$p \propto \frac{1}{V}$$

This equation can also be written as follows:

$$p = \frac{k}{V}$$

where k is a proportionality constant. If we rearrange this equation, we can say that:

$$pV = k$$

This equation means that, assuming the temperature is constant, multiplying any pressure and volume values for a fixed amount of gas will always give the same value. So, for example, $p_1V_1 = k$ and $p_2V_2 = k$, where the subscripts 1 and 2 refer to two pairs of pressure and volume readings for the same mass of gas at the same temperature.

From this, we can then say that:

$$p_1V_1 = p_2V_2$$

Important: Remember that Boyle's Law requires two conditions. First, the amount of gas must stay constant. Clearly, if you let a little of the air escape from the container in which it is enclosed, the pressure of the gas will decrease along with the volume, and the inverse proportion relationship is broken. Second, the temperature must stay constant. Cooling or heating matter generally causes it to contract or expand. In our original syringe demonstration, if you were to heat up the gas in the syringe, it would expand and force you to apply a greater force to keep the plunger at a given position. Again, the proportionality would be broken.

In the gas equations, k is a "variable constant". This means that k is constant in a particular set of situations, but in two different sets of situations it has different constant values.

Activity :: Investigation : Boyle's Law

Here are some of Boyle's original data. Note that pressure would originally have been measured using a *mercury manometer* and the units for pressure would have been *millimetres mercury* or mm Hg. However, to make things a bit easier for you, the pressure data have been converted to a unit that is more familiar. Note that the volume is given in terms of arbitrary marks (evenly made).

Volume (graduation mark)	**Pressure** (kPa)	**Volume** (graduation mark)	**Pressure** (kPa)
12	398	28	170
14	340	30	159
16	298	32	150
18	264	34	141
20	239	36	133
22	217	38	125
24	199	40	120
26	184		

1. Plot a graph of pressure (p) against volume (V). Volume will be on the x-axis and pressure on the y-axis. Describe the relationship that you see.

2. Plot a graph of p against $1/V$. Describe the relationship that you see.
3. Do your results support Boyle's Law? Explain your answer.

Interesting Fact

Did you know that the mechanisms involved in *breathing* also relate to Boyle's Law? Just below the lungs is a muscle called the **diaphragm**. When a person breathes in, the diaphragm moves down and becomes more 'flattened' so that the volume of the lungs can increase. When the lung volume *increases*, the pressure in the lungs *decreases* (Boyle's law). Since air always moves from areas of high pressure to areas of lower pressure, air will now be drawn into the lungs because the air pressure *outside* the body is higher than the pressure *in* the lungs. The opposite process happens when a person breathes out. Now, the diaphragm moves upwards and causes the volume of the lungs to *decrease*. The pressure in the lungs will *increase*, and the air that was in the lungs will be forced out towards the lower air pressure outside the body.

Worked Example 26: Boyle's Law 1

Question: A sample of helium occupies a volume of 160 cm^3 at 100 kPa and 25 °C. What volume will it occupy if the pressure is adjusted to 80 kPa and if the temperature remains unchanged?

Answer

Step 4 : Write down all the information that you know about the gas.

V_1 = 160 cm^3 and V_2 = ?

p_1 = 100 kPa and p_2 = 80 kPa

Step 1 : Use an appropriate gas law equation to calculate the unknown variable.

Because the temperature of the gas stays the same, the following equation can be used:

$$p_1V_1 = p_2V_2$$

If the equation is rearranged, then

$$V_2 = \frac{p_1V_1}{p_2}$$

Step 2 : Substitute the known values into the equation, making sure that the units for each variable are the same. Calculate the unknown variable.

$$V_2 = \frac{100 \times 160}{80} = 200cm^3$$

The volume occupied by the gas at a pressure of 80kPa, is 200 cm^3

Worked Example 27: Boyle's Law 2

Question: The pressure on a 2.5 l volume of gas is increased from 695 Pa to 755 Pa while a constant temperature is maintained. What is the volume of the gas under these pressure conditions?

Answer

Step 1 : Write down all the information that you know about the gas.

V_1 = 2.5 l and V_2 = ?

p_1 = 695 Pa and p_2 = 755 Pa

Step 2 : Choose a relevant gas law equation to calculate the unknown variable.

At constant temperature,

$$p_1V_1 = p_2V_2$$

Therefore,

$$V_2 = \frac{p_1V_1}{p_2}$$

Step 3 : Substitute the known values into the equation, making sure that the units for each variable are the same. Calculate the unknown variable.

$$V_2 = \frac{695 \times 2.5}{755} = 2.3l$$

Important:

It is not necessary to convert to Standard International (SI) units in the examples we have used above. Changing pressure and volume into different units involves *multiplication*. If you were to change the units in the above equation, this would involve multiplication on both sides of the equation, and so the conversions cancel each other out. However, although *SI* units don't have to be used, you must make sure that for each variable you use the *same* units throughout the equation. This is not true for some of the calculations we will do at a later stage, where SI units *must* be used.

Exercise: Boyle's Law

1. An unknown gas has an initial pressure of 150 kPa and a volume of 1 L. If the volume is increased to 1.5 L, what will the pressure now be?
2. A bicycle pump contains 250 cm^3 of air at a pressure of 90 kPa. If the air is compressed, the volume is reduced to 200 cm^3. What is the pressure of the air inside the pump?
3. The air inside a syringe occupies a volume of 10 cm^3 and exerts a pressure of 100 kPa. If the end of the syringe is sealed and the plunger is pushed down, the pressure increases to 120 kPa. What is the volume of the air in the syringe?
4. During an investigation to find the relationship between the pressure and volume of an enclosed gas at constant temperature, the following results were obtained.

Volume (cm^3)	Pressure (kPa)
40	125.0
30	166.7
25	200.0

(a) For the results given in the above table, plot a graph of **pressure** (y-axis) against the **inverse of volume** (x-axis).

(b) From the graph, deduce the relationship between the pressure and volume of an enclosed gas at constant temperature.

(c) Use the graph to predict what the volume of the gas would be at a pressure of 40 kPa. Show on your graph how you arrived at your answer.

(*IEB 2004 Paper 2*)

8.3 Charles's Law: Volume and Temperature of an enclosed gas

Charles's law describes the relationship between the **volume** and **temperature** of a gas. The law was first published by Joseph Louis Gay-Lussac in 1802, but he referenced unpublished work by Jacques Charles from around 1787. This law states that at constant pressure, the volume of a given mass of an ideal gas increases or decreases by the same factor as its temperature (in kelvin) increases or decreases. Another way of saying this is that temperature and volume are **directly proportional** (figure **??**).

Definition: Charles's Law
The volume of an enclosed sample of gas is directly proportional to its absolute temperature provided the pressure is kept constant.

Charles's Law is also known as Gay-Lussac's Law. This is because Charles did not publish his discovery, and it was rediscovered independently by another French Chemist Joseph Louis Gay-Lussac some years later.

Activity :: Demonstration : Charles's Law
Aim:
To demonstrate Charles's Law using simple materials.
Apparatus:
glass bottle (e.g. empty glass coke bottle), balloon, bunsen burner, retort stand
Method:

1. Place the balloon over the opening of the empty bottle.
2. Place the bottle on the retort stand over the bunsen burner and allow it to heat up. Observe what happens to the balloon. WARNING: Be careful when handling the heated bottle. You may need to wear gloves for protection.

Results:

You should see that the balloon starts to expand. As the air inside the bottle is heated, the pressure also increases, causing the volume to increase. Since the volume of the glass bottle can't increase, the air moves into the balloon, causing it to expand.

Conclusion:

The temperature and volume of the gas are directly related to each other. As one increases, so does the other.

Mathematically, the relationship between temperature and pressure can be represented as follows:

$$V \propto T$$

or

$$V = kT$$

If the equation is rearranged, then...

$$\frac{V}{T} = k$$

and, following the same logic that was used for Boyle's law:

$$\frac{V_1}{T_1} = \frac{V_2}{T_2}$$

The equation relating volume and temperature produces a straight line graph (refer back to the notes on proportionality if this is unclear). This relationship is shown in figure 8.3.

Figure 8.3: The volume of a gas is directly proportional to its temperature, provided the pressure of the gas is constant.

However, if this graph is plotted on a **celsius** temperature scale, the zero point of temperature doesn't correspond to the zero point of volume. When the volume is zero, the temperature is actually -273.15^{0}C (figure 8.4.

A new temperature scale, the **Kelvin scale** must be used instead. Since zero on the Celsius scale corresponds with a Kelvin temperature of -273.15^{0}C, it can be said that:

Kelvin temperature (T) = Celsius temperature (t) + 273.15

Figure 8.4: The relationship between volume and temperature, shown on a Celsius temperature scale.

At school level, you can simplify this slightly and convert between the two temperature scales as follows:

$$T = t + 273$$
or
$$t = T - 273$$

Can you explain Charles's law in terms of the kinetic theory of gases? When the temperature of a gas increases, so does the average speed of its molecules. The molecules collide with the walls of the container more often and with greater impact. These collisions will push back the walls, so that the gas occupies a greater volume than it did at the start. We saw this in the first demonstration. Because the glass bottle couldn't expand, the gas pushed out the balloon instead.

?

Exercise: Charles's law

The table below gives the temperature (in ^{0}C) of a number of gases under different volumes at a constant pressure.

Volume (l)	He	H_2	N_2O
0	-272.4	-271.8	-275.0
0.25	-245.5	-192.4	-123.5
0.5	-218.6	-113.1	28.1
0.75	-191.8	-33.7	179.6
1.0	-164.9	45.7	331.1
1.5	-111.1	204.4	634.1
2	-57.4	363.1	937.2
2.5	-3.6	521.8	1240.2
3.0	50.2	680.6	1543.2
3.5	103.9	839.3	1846.2

1. On the same set of axes, draw graphs to show the relationship between temperature and volume for each of the gases.
2. Describe the relationship you observe.
3. If you extrapolate the graphs (in other words, extend the graph line even though you may not have the exact data points), at what temperature do they intersect?
4. What is significant about this temperature?

Worked Example 28: Charles's Law 1

Question: Ammonium chloride and calcium hydroxide are allowed to react. The ammonia that is released in the reaction is collected in a gas syringe and sealed in. This gas is allowed to come to room temperature which is 32°C. The volume of the ammonia is found to be 122 ml. It is now placed in a water bath set at 7°C. What will be the volume reading after the syringe has been left in the bath for a good while (assume the plunger moves completely freely)?

Answer
Step 1 : Write down all the information that you know about the gas.
V_1 = 122 ml and V_2 = ?
T_1 = 32^0C and T_2 = 7^0C

Step 2 : Convert the known values to SI units if necessary.
Here, temperature must be converted into Kelvin, therefore:
T_1 = 32 + 273 = 305 K
T_2 = 7 + 273 = 280 K

Step 3 : Choose a relevant gas law equation that will allow you to calculate the unknown variable.

$$\frac{V_1}{T_1} = \frac{V_2}{T_2}$$

Therefore,

$$V_2 = \frac{V_1 \times T_2}{T_1}$$

Step 4 : Substitute the known values into the equation. Calculate the unknown variable.

$$V_2 = \frac{122 \times 280}{305} = 112ml$$

Important:
Note that here the temperature must be converted to Kelvin (SI) since the change from degrees Celcius involves addition, not multiplication by a fixed conversion ratio (as is the case with pressure and volume.)

Worked Example 29: Charles's Law 2

Question: At a temperature of 298 K, a certain amount of CO_2 gas occupies a volume of 6 l. What volume will the gas occupy if its temperature is reduced to 273 K?

Answer

Step 1 : Write down all the information that you know about the gas.
$V_1 = 6$ l and $V_2 = ?$
$T_1 = 298$ K and $T_2 = 273$ K

Step 2 : Convert the known values to SI units if necessary.
Temperature data is already in Kelvin, and so no conversions are necessary.

Step 3 : Choose a relevant gas law equation that will allow you to calculate the unknown variable.

$$\frac{V_1}{T_1} = \frac{V_2}{T_2}$$

Therefore,

$$V_2 = \frac{V_1 \times T_2}{T_1}$$

Step 4 : Substitute the known values into the equation. Calculate the unknown variable.

$$V_2 = \frac{6 \times 273}{298} = 5.5l$$

8.4 The relationship between temperature and pressure

The pressure of a gas is directly proportional to its temperature, if the volume is kept constant (figure 8.5). When the temperature of a gas increases, so does the energy of the particles. This causes them to move more rapidly and to collide with each other and with the side of the container more often. Since pressure is a measure of these collisions, the pressure of the gas increases with an increase in temperature. The pressure of the gas will decrease if its temperature decreases.

Figure 8.5: The relationship between the temperature and pressure of a gas

In the same way that we have done for the other gas laws, we can describe the relationship between temperature and pressure using symbols, as follows:

$$T \propto p, \text{ therefore } p = kT$$

We can also say that:

$$\frac{p}{T} = k$$

and that, provided the *amount* of gas stays the same...

$$\frac{p_1}{T_1} = \frac{p_2}{T_2}$$

Exercise: More gas laws

1. A gas of unknown volume has a temperature of 14°C. When the temperature of the gas is increased to 100°C, the volume is found to be 5.5 L. What was the initial volume of the gas?
2. A gas has an initial volume of 2600 mL and a temperature of 350 K.
   (a) If the volume is reduced to 1500 mL, what will the temperature of the gas be in Kelvin?
   (b) Has the temperature *increased* or *decreased*?
   (c) Explain this change, using the kinetic theory of matter.
3. A cylinder of propane gas at a temperature of 20°C exerts a pressure of 8 atm. When a cylinder has been placed in sunlight, its temperature increases to 25°C. What is the pressure of the gas inside the cylinder at this temperature?

8.5 The general gas equation

All the gas laws we have described so far rely on the fact that at least one variable (T, p or V) remains constant. Since this is unlikely to be the case most times, it is useful to combine the relationships into one equation. These relationships are as follows:

Boyle's law: p $\propto \frac{1}{V}$ (constant T)

Relationship between p and T: p $\propto$ T (constant V)

If we combine these relationships, we get p $\propto \frac{T}{V}$

If we introduce the proportionality constant k, we get p = k$\frac{T}{V}$

or, rearranging the equation...

$$pV = kT$$

We can also rewrite this relationship as follows:

$$\frac{pV}{T} = k$$

Provided the mass of the gas stays the same, we can also say that:

$$\frac{p_1V_1}{T_1} = \frac{p_2V_2}{T_2}$$

In the above equation, the subscripts 1 and 2 refer to two pressure and volume readings for the same mass of gas under different conditions. This is known as the **general gas equation**.

Temperature is always in kelvin and the units used for pressure and volume must be the same on both sides of the equation.

Important:
Remember that the general gas equation only applies if the mass of the gas is fixed.

Worked Example 30: General Gas Equation 1

Question: At the beginning of a journey, a truck tyre has a volume of 30 dm^3 and an internal pressure of 170 kPa. The temperature of the tyre is 16^0C. By the end of the trip, the volume of the tyre has increased to 32 dm^3 and the temperature of the air inside the tyre is 35^0C. What is the tyre pressure at the end of the journey?

Answer
Step 1 : Write down all the information that you know about the gas.
p_1 = 170 kPa and p_2 = ?
V_1 = 30 dm^3 and V_2 = 32 dm^3
T_1 = 16^0C and T_2 = 40^0C

Step 2 : Convert the known values to SI units if necessary.
Here, temperature must be converted into Kelvin, therefore:
T_1 = 16 + 273 = 289 K
T_2 = 40 + 273 = 313 K

Step 3 : Choose a relevant gas law equation that will allow you to calculate the unknown variable.
Use the general gas equation to solve this problem:

$$\frac{p_1 \times V_1}{T_1} = \frac{p_2 \times V_2}{T_2}$$

Therefore,

$$p_2 = \frac{p_1 \times V_1 \times T_2}{T_1 \times V_2}$$

Step 4 : Substitute the known values into the equation. Calculate the unknown variable.

$$p_2 = \frac{170 \times 30 \times 313}{289 \times 32} = 173kPa$$

The pressure of the tyre at the end of the journey is 173 kPa.

Worked Example 31: General Gas Equation 2

Question: A cylinder that contains methane gas is kept at a temperature of 15^0C and exerts a pressure of 7 atm. If the temperature of the cylinder increases to 25^0C, what pressure does the gas now exert? (Refer to table 8.1 to see what an 'atm' is.

Answer
Step 1 : Write down all the information that you know about the gas.
p_1 = 7 atm and p_2 = ?

$T_1 = 15^0C$ and $T_2 = 25^0C$

Step 2 : Convert the known values to SI units if necessary.
Here, temperature must be converted into Kelvin, therefore:
$T_1 = 15 + 273 = 288$ K
$T_2 = 25 + 273 = 298$ K

Step 3 : Choose a relevant gas law equation that will allow you to calculate the unknown variable.
Since the volume of the cylinder is constant, we can write:

$$\frac{p_1}{T_1} = \frac{p_2}{T_2}$$

Therefore,

$$p_2 = \frac{p_1 \times T_2}{T_1}$$

Step 4 : Substitute the known values into the equation. Calculate the unknown variable.

$$p_2 = \frac{7 \times 298}{288} = 7.24atm$$

The pressure of the gas is 7.24 atm.

Worked Example 32: General Gas Equation 3

Question: A gas container can withstand a pressure of 130 kPa before it will start to leak. Assuming that the volume of the gas in the container stays the same, at what temperature will the container start to leak if the gas exerts a pressure of 100 kPa at 15^0C?

Answer
Step 1 : Write down all the information that you know about the gas.
$p_1 = 100$ kPa and $p_2 = 130$ kPa
$T_1 = 15^0C$ and $T_2 = ?$

Step 2 : Convert the known values to SI units if necessary.
Here, temperature must be converted into Kelvin, therefore:
$T_1 = 15 + 273 = 288$ K

Step 3 : Choose a relevant gas law equation that will allow you to calculate the unknown variable.
Since the volume of the container is constant, we can write:

$$\frac{p_1}{T_1} = \frac{p_2}{T_2}$$

Therefore,

$$\frac{1}{T_2} = \frac{p_1}{T_1 \times p_2}$$

Therefore,

$$T_2 = \frac{T_1 \times p_2}{p_1}$$

Step 4 : Substitute the known values into the equation. Calculate the unknown variable.

$$T_2 = \frac{288 \times 130}{100} = 374.4K = 101.4^0C$$

Exercise: The general gas equation

1. A closed gas system initially has a volume of 8 L and a temperature of 100°C. The pressure of the gas is unknown. If the temperature of the gas decreases to 50°C, the gas occupies a volume of 5 L. If the pressure of the gas under these conditions is 1.2 atm, what was the initial pressure of the gas?
2. A balloon is filled with helium gas at 27°C and a pressure of 1.0 atm. As the balloon rises, the volume of the balloon increases by a factor of 1.6 and the temperature decreases to 15°C. What is the final pressure of the gas (assuming none has escaped)?
3. 25 cm^3 of gas at 1 atm has a temperature of 20°C. When the gas is compressed to 20 cm^3, the temperature of the gas increases to 28°C. Calculate the final pressure of the gas.

8.6 The ideal gas equation

In the early 1800's, Amedeo Avogadro hypothesised that if you have samples of different gases, of the same volume, at a fixed temperature and pressure, then the samples must contain the same number of freely moving particles (i.e. atoms or molecules).

Definition: Avogadro's Law
Equal volumes of gases, at the same temperature and pressure, contain the same number of molecules.

You will remember from an earlier section, that we combined different gas law equations to get *one* that included temperature, volume and pressure. In this equation, pV = kT, the value of k is different for different masses of gas. If we were to measure the amount of gas in moles, then k = nR, where n is the number of moles of gas and R is the universal gas constant. The value of R is 8.3143 $J.K^{-1}$, or for most calculations, 8.3 $J.K^{-1}$. So, if we replace k in the general gas equation, we get the following **ideal gas equation**.

$$pV = nRT$$

Important:

1. The value of R is the same for all gases
2. All quantities in the equation pV = nRT must be in the same units as the value of R. In other words, SI units must be used throughout the equation.

The following table may help you when you convert to SI units.

Table 8.1: Conversion table showing different units of measurement for volume, pressure and temperature.

Variable	Pressure (**p**)	Volume (**V**)	moles (**n**)	universal gas constant (**R**)	temperature (**K**)
SI unit	Pascals (Pa)	m^3	mol	$J.mol.K^{-1}$	kelvin (K)
Other units and conversions	760 mm Hg = 1 atm = 101325 Pa = 101.325 kPa	1 m^3 = 1000000 cm^3 = 1000 dm^3 = 1000 litres			K = ^{0}C + 273

Worked Example 33: Ideal gas equation 1

Question: Two moles of oxygen (O_2) gas occupy a volume of 25 dm^3 at a temperature of 40^{0}C. Calculate the pressure of the gas under these conditions.

Answer
Step 1 : Write down all the information that you know about the gas.
p = ?
V = 25 dm^3
n = 2
T = 40^{0}C

Step 2 : Convert the known values to SI units if necessary.

$$V = \frac{25}{1000} = 0.025m^3$$

$$T = 40 + 273 = 313K$$

Step 3 : Choose a relevant gas law equation that will allow you to calculate the unknown variable.

$$pV = nRT$$

Therefore,

$$p = \frac{nRT}{V}$$

Step 4 : Substitute the known values into the equation. Calculate the unknown variable.

$$p = 2 \times 8.3 \times 3130.025 = 207832Pa = 207.8kPa$$

Worked Example 34: Ideal gas equation 2

Question: Carbon dioxide (CO_2) gas is produced as a result of the reaction between calcium carbonate and hydrochloric acid. The gas that is produced is collected in a 20 dm^3 container. The pressure of the gas is 105 kPa at a temperature of 20^{0}C.

What mass of carbon dioxide was produced?

Answer
Step 1 : Write down all the information that you know about the gas.
p = 105 kPa
V = 20 dm^3
T = 20^0C

Step 2 : Convert the known values to SI units if necessary.

$$p = 105 \times 1000 = 105000Pa$$

$$T = 20 + 273 = 293K$$

$$V = \frac{20}{1000} = 0.02m^3$$

Step 3 : Choose a relevant gas law equation that will allow you to calculate the unknown variable.

$$pV = nRT$$

Therefore,

$$n = \frac{pV}{RT}$$

Step 4 : Substitute the known values into the equation. Calculate the unknown variable.

$$n = \frac{105000 \times 0.02}{8.3 \times 293} = 0.86moles$$

Step 5 : Calculate mass from moles

$$n = \frac{m}{M}$$

Therefore,

$$m = n \times M$$

The molar mass of CO_2 is calculated as follows:

$$M = 12 + (2 \times 16) = 44g.mol^{-1}$$

Therefore,

$$m = 0.86 \times 44 = 37.84g$$

Worked Example 35: Ideal gas equation 3

Question: 1 mole of nitrogen (N_2) reacts with hydrogen (H_2) according to the following equation:

$$N_2 + 3H_2 \rightarrow 2NH_3$$

The ammonia (NH_3) gas is collected in a separate gas cylinder which has a volume of 25 dm^3. The temperature of the gas is 22^0C. Calculate the pressure of the gas inside the cylinder.

Answer
Step 1 : Write down all the information that you know about the gas.
V = 25 dm^3
n = 2 (Calculate this by looking at the mole ratio of nitrogen to ammonia, which is 1:2)
T = 22^0C

Step 2 : Convert the known values to SI units if necessary.

$$V = \frac{25}{1000} = 0.025m^3$$

$$T = 22 + 273 = 295K$$

Step 3 : Choose a relevant gas law equation that will allow you to calculate the unknown variable.

$$pV = nRT$$

Therefore,

$$p = \frac{nRT}{V}$$

Step 4 : Substitute the known values into the equation. Calculate the unknown variable.

$$p = \frac{2 \times 8.3 \times 295}{0.025} = 195880Pa = 195.89kPa$$

Worked Example 36: Ideal gas equation 4

Question: Calculate the number of air particles in a 10 m by 7 m by 2 m classroom on a day when the temperature is 23°C and the air pressure is 98 kPa.

Answer
Step 1 : Write down all the information that you know about the gas.
V = 10 m × 7 m × 2m = 140 m^3
p = 98 kPa
T = 23^0C

Step 2 : Convert the known values to SI units if necessary.

$$p = 98 \times 1000 = 98000Pa$$

$$T = 23 + 273 = 296K$$

Step 3 : Choose a relevant gas law equation that will allow you to calculate the unknown variable.

$$pV = nRT$$

Therefore,

$$n = \frac{pV}{RT}$$

Step 4 : Substitute the known values into the equation. Calculate the unknown variable.

$$n = \frac{98000 \times 140}{8.3 \times 296} = 5584.5mol$$

Worked Example 37: Applying the gas laws

Question: Most modern cars are equipped with airbags for both the driver and the passenger. An airbag will completely inflate in 0,05 s. This is important because a typical car collision lasts about 0,125 s. The following reaction of sodium azide (a compound found in airbags) is activated by an electrical signal:

$$2NaN_3(s) \rightarrow 2Na(s) + 3N_2(g)$$

1. Calculate the mass of $N_2(g)$ needed to inflate a sample airbag to a volume of 65 dm^3 at 25 °C and 99,3 kPa. Assume the gas temperature remains constant during the reaction.
2. In reality the above reaction is exothermic. Describe, in terms of the kinetic molecular theory, how the pressure in the sample airbag will change, if at all, as the gas temperature returns to 25 °C.

Answer

Step 1 : Look at the information you have been given, and the information you still need.

Here you are given the volume, temperature and pressure. You are required to work out the mass of N_2.

Step 2 : Check that all the units are S.I. units

Pressure: $93{,}3 \times 10^3$ Pa
Volume: 65×10^{-3} m^3
Temperature: $(273 + 25)$ K
Gas Constant: $8{,}31$

Step 3 : Write out the Ideal Gas formula

$$pV = nRT$$

Step 4 : Solve for the required quantity using symbols

$$n = \frac{pV}{RT}$$

Step 5 : Solve by substituting numbers into the equation to solve for 'n'.

$$n = \frac{99{,}3 \times 10^3 \times 65 \times 10^{-3}}{8{,}31 \times (273 + 25)}$$

Step 6 : Convert the number of moles to number of grams

$$\begin{aligned} m &= n \times M \\ m &= 2{,}61 \times 28 \\ m &= 73{,}0g \end{aligned}$$

Step 7 : Theory Question
When the temperature decreases the intensity of collisions with the walls of the airbag and between particles decreases. Therefore pressure decreases.

Exercise: The ideal gas equation

1. An unknown gas has pressure, volume and temperature of 0.9 atm, 8 L and 120°C respectively. How many moles of gas are present?
2. 6 g of chlorine (Cl_2) occupies a volume of 0.002 m^3 at a temperature of 26°C. What is the pressure of the gas under these conditions?
3. An average pair of human lungs contains about 3.5 L of air after inhalation and about 3.0 L after exhalation. Assuming that air in your lungs is at 37°C and 1.0 atm, determine the number of moles of air in a typical breath.
4. A learner is asked to calculate the answer to the problem below:
   Calculate the pressure exerted by 1.5 moles of nitrogen gas in a container with a volume of 20 dm^3 at a temperature of 37°C.
   The learner writes the solution as follows:
   V = 20 dm^3
   n = 1.5 mol
   R = 8.3 $J.K^{-1}.mol^{-1}$
   T = 37 + 273 = 310 K
   p = nRT, therefore

$$p = \frac{nRV}{T}$$

$$= \frac{1.5 \times 8.3 \times 20}{310}$$

   = 0.8 kPa

   (a) Identify 2 mistakes the learner has made in the calculation.
   (b) Are the units of the final answer correct?
   (c) Rewrite the solution, correcting the mistakes to arrive at the right answer.

8.7 Molar volume of gases

It is possible to calculate the volume of a mole of gas at STP using what we now know about gases.

1. *Write down the ideal gas equation*

$$pV = nRT, \text{ therefore } V = \frac{nRT}{p}$$

2. *Record the values that you know, making sure that they are in SI units*

   You know that the gas is under STP conditions. These are as follows:

   p = 101.3 kPa = 101300 Pa

   n = 1 mole

   R = 8.3 $J.K^{-1}.mol^{-1}$

   T = 273 K

3. *Substitute these values into the original equation.*

$$V = \frac{nRT}{p}$$

$$V = \frac{1mol \times 8.3J.K^{-1}.mol^{-1} \times 273K}{101300Pa}$$

4. *Calculate the volume of 1 mole of gas under these conditions*

   The volume of 1 mole of gas at STP is 22.4 $\times$ 10^{-3} m^3 = 22.4 dm^3.

8.8 Ideal gases and non-ideal gas behaviour

In looking at the behaviour of gases to arrive at the Ideal Gas Law, we have limited our examination to a small range of temperature and pressure. Most gases do obey these laws most of the time, and are called **ideal gases**, but there are deviations at **high pressures** and **low temperatures**. So what is happening at these two extremes?

Earlier when we discussed the kinetic theory of gases, we made a number of assumptions about the behaviour of gases. We now need to look at two of these again because they affect how gases behave either when pressures are high or when temperatures are low.

1. *Molecules do occupy volume*

   This means that when pressures are very high and the molecules are compressed, their volume becomes significant. This means that the total volume available for the gas molecules to move is reduced and collisions become more frequent. This causes the pressure of the gas to be *higher* than what would normally have been predicted by Boyle's law (figure 8.6).

Figure 8.6: Gases deviate from ideal gas behaviour at high pressure.

2. *Forces of attraction do exist between molecules*

   At low temperatures, when the speed of the molecules decreases and they move closer together, the intermolecular forces become more apparent. As the attraction between molecules increases, their movement decreases and there are fewer collisions between them. The pressure of the gas at low temperatures is therefore lower than what would have been expected for an ideal gas (figure 8.7. If the temperature is low enough or the pressure high enough, a real gas will **liquify**.

Figure 8.7: Gases deviate from ideal gas behaviour at low temperatures

8.9 Summary

- The **kinetic theory of matter** helps to explain the behaviour of gases under different conditions.
- An **ideal gas** is one that obeys all the assumptions of the kinetic theory.
- A **real gas** behaves like an ideal gas, except at high pressures and low temperatures. Under these conditions, the forces between molecules become significant and the gas will liquify.
- **Boyle's law** states that the pressure of a fixed quantity of gas is inversely proportional to its volume, as long as the temperature stays the same. In other words, pV = k or

$$p_1V_1 = p_2V_2.$$

- **Charles's law** states that the volume of an enclosed sample of gas is directly proportional to its temperature, as long as the pressure stays the same. In other words,

$$\frac{V_1}{T_1} = \frac{V_2}{T_2}$$

- The **temperature** of a fixed mass of gas is directly proportional to its pressure, if the volume is constant. In other words,

$$\frac{p_1}{T_1} = \frac{p_2}{T_2}$$

- In the above equations, temperature must be written in **Kelvin**. Temperature in degrees Celsius (temperature = t) can be converted to temperature in Kelvin (temperature = T) using the following equation:

$$T = t + 273$$

- Combining Boyle's law and the relationship between the temperature and pressure of a gas, gives the **general gas equation**, which applies as long as the amount of gas remains constant. The general gas equation is pV = kT, or

$$\frac{p_1V_1}{T_1} = \frac{p_2V_2}{T_2}$$

- Because the mass of gas is not always constant, another equation is needed for these situations. The **ideal gas equation** can be written as

$$pV = nRT$$

where n is the number of moles of gas and R is the universal gas constant, which is 8.3 J.K^{-1}.mol^{-1}. In this equation, **SI units** must be used. Volume (m^3), pressure (Pa) and temperature (K).

- The **volume** of one mole of gas under STP is 22.4 dm^3. This is called the **molar gas volume**.

s

Exercise: Summary exercise

1. For each of the following, say whether the statement is **true** or **false**. If the statement is false, rewrite the statement correctly.
   (a) Real gases behave like ideal gases, except at low pressures and low temperatures.
   (b) The volume of a given mass of gas is inversely proportional to the pressure it exerts.
   (c) The temperature of a fixed mass of gas is directly proportional to its pressure, regardless of the volume of the gas.
2. For each of the following multiple choice questions, choose the **one correct answer**.
   (a) Which one of the following properties of a fixed quantity of a gas must be kept constant during a Boyle's law investigation?
      i. density
      ii. pressure
      iii. temperature
      iv. volume

      (*IEB 2003 Paper 2*)
   (b) Three containers of EQUAL VOLUME are filled with EQUAL MASSES of helium, nitrogen and carbon dioxide gas respectively. The gases in the three containers are all at the same TEMPERATURE. Which one of the following statements is correct regarding the pressure of the gases?
      i. All three gases will be at the same pressure
      ii. The helium will be at the greatest pressure
      iii. The nitrogen will be at the greatest pressure
      iv. The carbon dioxide will be at the greatest pressure

      (*IEB 2004 Paper 2*)
   (c) One mole of an ideal gas is stored at a temperature T (in Kelvin) in a rigid gas tank. If the average speed of the gas particles is doubled, what is the new Kelvin temperature of the gas?
      i. 4T
      ii. 2T
      iii. $\sqrt{2}$T

iv. 0.5 T

(*IEB 2002 Paper 2*)

(d) The ideal gas equation is given by **pV = nRT**. Which one of the following conditions is true according to Avogadro's hypothesis?

a	$p \propto 1/V$	(T = constant)
b	$V \propto T$	(p = constant)
c	$V \propto n$	(p, T = constant)
d	$p \propto T$	(n = constant)

(*DoE Exemplar paper 2, 2007*)

3. Use your knowledge of the gas laws to explain the following statements.
   (a) It is dangerous to put an aerosol can near heat.
   (b) A pressure vessel that is poorly designed and made can be a serious safety hazard (a pressure vessel is a closed, rigi container that is used to hold gases at a pressure that is higher than the normal air pressure).
   (c) The volume of a car tyre increases after a trip on a hot road.
4. Copy the following set of labelled axes and answer the questions that follow:

   (a) On the axes, **using a solid line**, draw the graph that would be obtained for a fixed mass of an ideal gas if the pressure is kept constant.
   (b) If the gradient of the above graph is measured to be 0.008 $m^3.K^{-1}$, calculate the pressure that 0.3 mol of this gas would exert.

   (**IEB 2002 Paper 2**)
5. Two gas cylinders, A and B, have a volume of 0.15 m^3 and 0.20 m^3 respectively. Cylinder A contains 1.25 mol He gas at pressure p and cylinder B contains 2.45 mol He gas at standard pressure. The ratio of the Kelvin temperatures A:B is 1.80:1.00. Calculate the pressure of the gas (in kPa) in cylinder A.

   (*IEB 2002 Paper 2*)
6. A learner investigates the relationship between the Celsius temperature and the pressure of a fixed amount of helium gas in a 500 cm^3 closed container. From the results of the investigation, she draws the graph below:

   (a) Under the conditions of this investigation, helium gas behaves like an ideal gas. Explain briefly why this is so.

(b) From the shape of the graph, the learner concludes that the pressure of the helium gas is directly proportional to the Celcius temperature. Is her conclusion correct? Briefly explain your answer.

(c) Calculate the pressure of the helium gas at 0 °C.

(d) Calculate the mass of helium gas in the container.

(*IEB 2003 Paper 2*)

7. One of the cylinders of a motor car engine, before compression contains 450 cm^3 of a mixture of air and petrol in the gaseous phase, at a temperature of 30°C and a pressure of 100 kPa. If the volume of the cylinder after compression decreases to one tenth of the original volume, and the temperature of the gas mixture rises to 140°C, calculate the pressure now exerted by the gas mixture.

8. In an experiment to determine the relationship between pressure and temperature of a fixed mass of gas, a group of learners obtained the following results:

Pressure (kPa)	101	120	130.5	138
Temperature (^{0}C)	0	50	80	100
Total gas volume (cm^3)	250	250	250	250

(a) Draw a straight-line graph of pressure (on the dependent, y-axis) versus temperature (on the independent, x-axis) on a piece of graph paper. Plot the points. Give your graph a suitable heading.
A straight-line graph passing through the origin is essential to obtain a mathematical relationship between pressure and temperature.

(b) Extrapolate (extend) your graph and determine the temperature (in ^{0}C) at which the graph will pass through the temperature axis.

(c) Write down, in words, the relationship between pressure and Kelvin temperature.

(d) From your graph, determine the pressure (in kPa) at 173 K. Indicate on your graph how you obtained this value.

(e) How would the gradient of the graph be affected (if at all) if a larger mass of the gas is used? Write down ONLY **increases**, **decreases** or **stays the same**.

(*DoE Exemplar Paper 2, 2007*)

Chapter 9

Organic Molecules - Grade 12

9.1 What is organic chemistry?

Organic chemistry is the branch of chemistry that deals with **organic molecules**. An organic molecule is one which contains **carbon**, and these molecules can range in size from simple molecules to complex structures containing thousands of atoms! Although the main element in organic compounds is carbon, other elements such as hydrogen (H), oxygen (O), nitrogen (N), sulfur (S) and phosphorus (P) are also common in these molecules.

Until the early nineteenth century, chemists had managed to make many simple compounds in the laboratory, but were still unable to produce the complex molecules that they found in living organisms. It was around this time that a Swedish chemist called **Jons Jakob Berzelius** suggested that compounds found only in living organisms (the organic compounds) should be grouped separately from those found in the non-living world (the inorganic compounds). He also suggested that the laws that governed how organic compounds formed, were different from those for inorganic compounds. From this, the idea developed that there was a 'vital force' in organic compounds. In other words, scientists believed that organic compounds would not follow the normal physical and chemical laws that applied to other inorganic compounds because the very 'force of life' made them different.

This idea of a mystical 'vital force' in organic compounds was weakened when scientists began to manufacture organic compounds in the laboratory from non-living materials. One of the first to do this was **Friedrich Wohler** in 1828, who successfully prepared urea, an organic compound in the urine of animals which, until that point, had only been found in animals. A few years later a student of Wohler's, **Hermann Kolbe**, made the organic compound *acetic acid* from inorganic compounds. By this stage it was acknowledged that organic compounds are governed by exactly the same laws that apply to inorganic compounds. The properties of organic compounds are not due to a 'vital force' but to the unique properties of the carbon atom itself.

Organic compounds are very important in daily life. They make up a big part of our own bodies, they are in the food we eat and in the clothes we wear. Organic compounds are also used to make products such as medicines, plastics, washing powders, dyes, along with a list of other items.

9.2 Sources of carbon

The main source of the carbon in organic compounds is **carbon dioxide** in the air. Plants use sunlight to convert carbon dioxide into organic compounds through the process of **photosynthesis**. Plants are therefore able to make their own organic compounds through photosynthesis, while animals feed on plants or plant products so that they gain the organic compounds that they need to survive.

Another important source of carbon is **fossil fuels** such as coal, petroleum and natural gas. This is because fossil fuels are themselves formed from the decaying remains of dead organisms (refer to chapter 21 for more information on fossil fuels).

9.3 Unique properties of carbon

Carbon has a number of unique properties which influence how it behaves and how it bonds with other atoms:

- Carbon has *four valence electrons* which means that each carbon atom can form bonds with four other atoms. Because of this, long *chain structures* can form. These chains can either be *unbranched* (figure 9.1) or *branched* (figure 9.2). Because of the number of bonds that carbon can form with other atoms, organic compounds can be very complex.

```
   |    |    |    |
— C — C — C — C —
   |    |    |    |
```

Figure 9.1: An unbranched carbon chain

```
   |    |    |    |
— C — C — C — C —
   |    |    |    |
     — C —
        |
     — C —
        |
     — C —
        |
```

Figure 9.2: A branched carbon chain

- Because of its position on the Periodic Table, most of the bonds that carbon forms with other atoms are *covalent*. Think for example of a C-C bond. The difference in electronegativity between the two atoms is zero, so this is a pure covalent bond. In the case of a C-H bond, the difference in electronegativity between carbon (2.5) and hydrogen (2.1) is so small that C-H bonds are almost purely covalent. The result of this is that most organic compounds are non-polar. This affects some of the properties of organic compounds.

9.4 Representing organic compounds

There are a number of ways to represent organic compounds. It is useful to know all of these so that you can recognise a molecule however it is shown. There are three main ways of representing a compound. We will use the example of a molecule called 2-methylpropane to help explain the difference between each.

9.4.1 Molecular formula

The molecular formula of a compound shows how many atoms of each type are in a molecule. The number of each atom is written as a subscript after the atomic symbol. The molecular formula of 2-methylpropane is:

$$C_4H_{10}$$

9.4.2 Structural formula

The structural formula of an organic compound shows every bond between every atom in the molecule. Each bond is represented by a line. The structural formula of 2-methylpropane is shown in figure 9.3.

Figure 9.3: The structural formula of 2-methylpropane

9.4.3 Condensed structural formula

When a compound is represented using its condensed structural formula, each carbon atom and the hydrogen atoms that are bonded directly to it are listed as a molecular formula, followed by a similar molecular formula for the neighbouring carbon atom. Branched groups are shown in brackets after the carbon atom to which they are bonded. The condensed structural formula below shows that in 2-methylpropane, there is a branched chain attached to the second carbon atom of the main chain. You can check this by looking at the structural formula in figure **??**.

$$CH_3CH(CH_3)CH_3$$

Exercise: Representing organic compounds

1. For each of the following organic compounds, give the **condensed structural formula** and the **molecular formula**.

(a)

(b)

2. For each of the following, give the **structural formula** and the **molecular formula**.
   (a) $CH_3CH_2CH_3$
   (b) $CH_3CH_2CH(CH_3)CH_3$
   (c) C_2H_6
3. Give two possible structural formulae for the compound with a molecular formula of C_4H_{10}.

9.5 Isomerism in organic compounds

It is possible for two organic compounds to have the same *molecular formula* but a different *structural formula*. Look for example at the two organic compounds that are shown in figure 9.4.

Figure 9.4: Isomers of a 4-carbon organic compound

If you were to count the number of carbon and hydrogen atoms in each compound, you would find that they are the same. They both have the same molecular formula (C_4H_{10}), but their structure is different and so are their properties. Such compounds are called **isomers**.

Definition: Isomer
In chemistry, isomers are molecules with the same molecular formula and often with the same kinds of chemical bonds between atoms, but in which the atoms are arranged differently.

Exercise: Isomers
Match the organic compound in Column A with its isomer Column B:

Column A	Column B
$CH_3CH(CH_3)OH$	$CH_3CH(CH_3)CH_3$
	C_3H_7OH

9.6 Functional groups

All organic compounds have a particular bond or group of atoms which we call its **functional group**. This group is important in determining how a compound will react.

Definition: Functional group
In organic chemistry, a functional group is a specific group of atoms within molecules, that are responsible for the characteristic chemical reactions of those molecules. The same functional group will undergo the same or similar chemical reaction(s) regardless of the size of the molecule it is a part of.

In one group of organic compounds called the **hydrocarbons**, the single, double and triple bonds of the alkanes, elkenes and alkynes are examples of functional groups. In another group, the alcohols, an oxygen and a hydrogen atom that are bonded to each other form the functional group for those compounds. All alcohols will contain an oxygen and a hydrogen atom bonded together in some part of the molecule.

Table 9.1 summarises some of the common functional groups. We will look at these in more detail later in this chapter.

9.7 The Hydrocarbons

Let us first look at a group of organic compounds known as the **hydrocarbons**. These molecules only contain carbon and hydrogen. The hydrocarbons that we are going to look at are called **aliphatic compounds**. The aliphatic compounds are divided into *acyclic compounds* (chain structures) and *cyclic compounds* (ring structures). The chain structures are further divided into structures that contain only *single bonds* (**alkanes**), those that contain *double bonds* (**alkenes**) and those that contain *triple bonds* (**alkynes**). Cyclic compounds include structures such as the *benzene ring*. Figure 9.5 summarises the classification of the hydrocarbons.

Hydrocarbons that contain only single bonds are called **saturated** hydrocarbons because each carbon atom is bonded to as many hydrogen atoms as possible. Figure 9.6 shows a molecule of ethane which is a saturated hydrocarbon.

Name of group	Functional group	Example	Diagram
Alk*ane*		Ethane	
Alk*ene*		Ethene	
Alk*yne*		Ethyne (acetylene)	
Halo-alkane	(X=F,Cl,Br,I)	Chloroethane	
Alcoh*ol*/ alkan*ol*		Ethanol	
Carboxylic acid		ethanoic acid	
Amine		Glycine	

Table 9.1: Some functional groups of organic compounds

Hydrocarbons that contain double or triple bonds are called **unsaturated** hydrocarbons because they don't contain as many hydrogen atoms as possible. Figure 9.7 shows a molecule of ethene which is an unsaturated hydrocarbon. If you compare the number of carbon and hydrogen atoms

Figure 9.5: The classification of the aliphatic hydrocarbons

Figure 9.6: A saturated hydrocarbon

in a molecule of ethane and a molecule of ethene, you will see that the number of hydrogen atoms in ethene is *less* than the number of hydrogen atoms in ethane despite the fact that they both contain two carbon atoms. In order for an unsaturated compound to become saturated, a double bond has to be broken, and another two hydrogen atoms added for each double bond that is replaced by a single bond.

Figure 9.7: An unsaturated hydrocarbon

Interesting Fact

Fat that occurs naturally in living matter such as animals and plants is used as food for human consumption and contains varying proportions of saturated and unsaturated fat. Foods that contain a high proportion of saturated fat are butter, ghee, suet, tallow, lard, coconut oil, cottonseed oil, and palm kernel oil, dairy products (especially cream and cheese), meat, and some prepared foods. Diets high in saturated fat are correlated with an increased incidence of atherosclerosis and coronary heart disease according to a number of studies. Vegetable oils contain unsaturated fats and can be hardened to form margarine by adding hydrogen on to some of the carbon=carbon double bonds using a nickel catalyst. The process is called hydrogenation

We will now go on to look at each of the hydrocarbon groups in more detail. These groups are the alkanes, the alkenes and the alkynes.

9.7.1 The Alkanes

The alkanes are hydrocarbons that only contain *single covalent bonds* between their carbon atoms. This means that they are *saturated* compounds and are quite unreactive. The simplest alkane has only one carbon atom and is called **methane**. This molecule is shown in figure 9.8.

Figure 9.8: The structural (a) and molecular formula (b) for methane

The second alkane in the series has two carbon atoms and is called **ethane**. This is shown in figure 9.9.

Figure 9.9: The structural (a) and molecular formula (b) for ethane

The third alkane in the series has three carbon atoms and is called **propane** (Figure 9.10).

Figure 9.10: The structural (a) and molecular formula (b) for propane

When you look at the molecular formula for each of the alkanes, you should notice a pattern developing. For each carbon atom that is added to the molecule, two hydrogen atoms are added. In other words, each molecule differs from the one before it by CH_2. This is called a *homologous series*. The alkanes have the general formula C_nH_{2n+2}.

The alkanes are the most important source of fuel in the world and are used extensively in the chemical industry. Some are gases (e.g. methane and ethane), while others are liquid fuels (e.g. octane, an important component of petrol).

Interesting Fact

Some fungi use alkanes as a source of carbon and energy. One fungus *Amorphotheca resinae* prefers the alkanes used in aviation fuel, and this can cause problems for aircraft in tropical areas!

9.7.2 Naming the alkanes

In order to give compounds a name, certain rules must be followed. When naming organic compounds, the IUPAC (International Union of Pure and Applied Chemistry) nomenclature is used. We will first look at some of the steps that need to be followed when naming a compound, and then try to apply these rules to some specific examples.

1. STEP 1: Recognise the *functional group* in the compound. This will determine the suffix (the 'end') of the name. For example, if the compound is an alkane, the suffix will be -ane; if the compound is an alkene the suffix will be -ene; if the compound is an alcohol the suffix will be -ol, and so on.
2. STEP 2: Find the longest continuous carbon chain (it won't always be a *straight* chain) and count the number of carbon atoms in this chain. This number will determine the prefix (the 'beginning') of the compound's name. These prefixes are shown in table 9.2. So, for example, an alkane that has 3 carbon atoms will have the suffix *prop* and the compound's name will be *propane*.

Carbon atoms	prefix
1	meth(ane)
2	eth(ane)
3	prop(ane)
4	but(ane)
5	pent(ane)
6	hex(ane)
7	hept(ane)
8	oct(ane)
9	non(ane)
10	dec(ane)

Table 9.2: The prefix of a compound's name is determined by the number of carbon atoms in the longest chain

3. STEP 3: Number the carbons in the longest carbon chain (Important: If there is a double or triple bond, you need to start numbering so that the bond is at the carbon with the lowest number.
4. STEP 4: Look for any branched groups and name them. Also give them a number to show their position on the carbon chain. If there are no branched groups, this step can be left out.
5. STEP 5: Combine the elements of the name into a single word in the following order: branched groups; prefix; name ending according to the functional group and its position along the longest carbon chain.

Worked Example 38: Naming the alkanes

Question: Give the IUPAC name for the following compound:
Note: The numbers attached to the carbon atoms would not normally be shown. The atoms have been numbered to help you to name the compound.

Answer
Step 1 : Identify the functional group
The compound is a hydrocarbon with single bonds between the carbon atoms. It is an alkane and will have a suffix of -ane.

Step 2 : Find the longest carbon chain

There are four carbon atoms in the longest chain. The prefix of the compound will be 'but'.

Step 3 : Number the carbons in the longest chain
In this case, it is easy. The carbons are numbered from left to right, from one to four.

Step 4 : Look for any branched groups, name them and give their position on the carbon chain
There are no branched groups in this compound.

Step 5 : Combine the elements of the name into a single word
The name of the compound is **butane**.

Worked Example 39: Naming the alkanes

Question: Give the IUPAC name for the following compound:

Answer
Step 1 : Identify the functional group
The compound is an alkane and will have the suffix -ane.

Step 2 : Find the longest carbon chain
There are three carbons in the longest chain. The prefix for this compound is -prop.

Step 3 : Number the carbons in the carbon chain
If we start at the carbon on the left, we can number the atoms as shown below:

Step 4 : Look for any branched groups, name them and give their position on the carbon chain
There is a branched group attached to the second carbon atom. This group has the formula CH_3 which is methane. However, because it is not part of the main chain, it is given the suffix -yl (i.e. methyl). The position of the methyl group comes just

before its name (see next step).

Step 5 : Combine the elements of the compound's name into a single word in the order of branched groups; prefix; name ending according to the functional group.
The compound's name is **2-methylpropane**.

Worked Example 40: Naming the alkanes

Question: Give the IUPAC name for the following compound:

$$CH_3CH(CH_3)CH(CH_3)CH_3$$

(Remember that the side groups are shown in brackets after the carbon atom to which they are attached.)
Answer
Step 1 : Draw the compound from its condensed structural formula
The structural formula of the compound is:

Step 2 : Identify the functional group
The compound is an alkane and will have the suffix -ane.

Step 3 : Find the longest carbon chain
There are four carbons in the longest chain. The prefix for this compound is -but.

Step 4 : Number the carbons in the carbon chain
If we start at the carbon on the left, carbon atoms are numbered as shown in the diagram above. A second way that the carbons could be numbered is:

Step 5 : Look for any branched groups, name them and give their position on the carbon chain
There are two methyl groups attached to the main chain. The first one is attached to the second carbon atom and the second methyl group is attached to the third carbon atom. Notice that in this example it does not matter how you have chosen to number the carbons in the main chain; the methyl groups are still attached to the second and third carbons and so the naming of the compound is not affected.

Step 6 : Combine the elements of the compound's name into a single word in the order of branched groups; prefix; name ending according to the functional group.
The compound's name is **2,3-dimethyl-butane**.

Worked Example 41: Naming the alkanes

Question: Give the IUPAC name for the following compound:

Answer
Step 1 : Identify the functional group
The compound is an alkane and will have the suffix -ane.

Step 2 : Find the longest carbon chain and number the carbons in the longest chain.
There are five carbons in the longest chain if they are numbered as shown below. The prefix for the compound is -pent.

Step 3 : Look for any branched groups, name them and give their position on the carbon chain
There are two methyl groups attached to the main chain. The first one is attached to the first carbon atom and the second methyl group is attached to the third carbon atom.

Step 4 : Combine the elements of the compound's name into a single word in the order of branched groups; prefix; name ending according to the functional group.
The compound's name is **1,3-dimethyl-pentane**.

Exercise: Naming the alkanes

1. Give the structural formula for each of the following:
   (a) Octane
   (b) $CH_3CH_2CH_3$
   (c) $CH_3CH(CH_3)CH_3$
   (d) 3-ethyl-pentane
2. Give the IUPAC name for each of the following organic compounds.

```
        H     H    CH3
        |     |     |
   H — C — C — C — H
        |     |     |
        H     H     H
```

   (a)
   (b) $CH_3CH_2CH(CH_3)CH_2CH_3$
   (c) $CH_3CH(CH_3)CH_2CH(CH_3)CH_3$

9.7.3 Properties of the alkanes

We have already mentioned that the alkanes are relatively unreactive because of their stable C-C and C-H bonds. The boiling point and melting point of these molecules is determined by their molecular structure, and their surface area. The more carbon atoms there are in an alkane, the greater the surface area and therefore the higher the boiling point. The melting point also increases as the number of carbon atoms in the molecule increases. This can be seen in the data in table 9.3.

Formula	Name	Melting point (^{0}C)	Boiling point (^{0}C)	Phase at room temperature
CH_4	methane	-183	-162	gas
C_2H_6	ethane	-182	-88	gas
C_3H_8	propane	-187	-45	gas
C_4H_{10}	butane	-138	-0.5	gas
C_5H_{12}	pentane	-130	36	liquid
C_6H_{14}	hexane	-95	69	liquid
$C_{17}H_{36}$	heptadecane	22	302	solid

Table 9.3: Properties of some of the alkanes

You will also notice that, when the molecular mass of the alkanes is low (i.e. there are few carbon atoms), the organic compounds are *gases* because the intermolecular forces are weak. As the number of carbon atoms and the molecular mass increases, the compounds are more likely to be liquids or solids because the intermolecular forces are stronger.

9.7.4 Reactions of the alkanes

There are three types of reactions that can occur in saturated compounds such as the alkanes.

1. **Substitution reactions**

   Substitution reactions involve the removal of a hydrogen atom which is replaced by an atom of another element, such as a halogen (F, Cl, Br or I) (figure 9.11). The product is called a **halo-alkane**. Since alkanes are not very reactive, heat or light are needed for this

Figure 9.11: A substitution reaction

reaction to take place.

e.g. $CH_2{=}CH_2 + HBr \rightarrow CH_3\text{-}CH_2\text{-}Br$ (halo-alkane)

Halo-alkanes (also sometimes called *alkyl halides*) that contain methane and chlorine are substances that can be used as anaesthetics during operations. One example is trichloromethane, also known as 'chloroform' (figure 9.12).

Figure 9.12: Trichloromethane

2. **Elimination reactions**

   Saturated compounds can also undergo elimination reactions to become unsaturated (figure 9.13). In the example below, an atom of hydrogen and chlorine are eliminated from the original compound to form an unsaturated halo-alkene.

   e.g. $CH_2Cl - CH_2Cl \rightarrow CH_2 = CHCl + HCl$

Figure 9.13: An elimination reaction

3. **Oxidation reactions**

   When alkanes are burnt in air, they react with the oxygen in air and heat is produced. This is called an oxidation or combustion reaction. Carbon dioxide and water are given off as products. Heat is also released during the reaction. The burning of alkanes provides most of the energy that is used by man.

   e.g. $CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O + heat$

Exercise: The Alkanes

1. Give the IUPAC name for each of the following alkanes:
   (a) C_6H_{14}

   (b)
   (c) CH_3CH_3
2. Give the structural formula for each of the following compounds:
   (a) octane
   (b) 3-methyl-hexane
3. Methane is one of the simplest alkanes and yet it is an important fuel source. Methane occurs naturally in wetlands, natural gas and permafrost. However, methane can also be produced when organic wastes (e.g. animal manure and decaying material) are broken down by bacteria under conditions that are anaerobic (there is no oxygen). The simplified reaction is shown below:

   $$\text{Organic matter} \rightarrow \text{Simple organic acids} \rightarrow \text{Biogas}$$

   The organic matter could be carbohydrates, proteins or fats which are broken down by acid-forming bacteria into simple organic acids such as acetic acid or formic acid. Methane-forming bacteria then convert these acids into biogases such as methane and ammonia.
   The production of methane in this way is very important because methane can be used as a fuel source. One of the advantages of methane over other fuels like coal, is that it produces more energy but with lower carbon dioxide emissions. The problem however, is that methane itself is a greenhouse gas and has a much higher global warming potential than carbon dioxide. So, producing methane may in fact have an even more dangerous impact on the environment.
   (a) What is the structural formula of methane?
   (b) Write an equation to show the reaction that takes place when methane is burned as a fuel.
   (c) Explain what is meant by the statement that methane 'has a greater global warming potential than carbon dioxide'.
4. Chlorine and ethane react to form chloroethane and hydrogen chloride.
   (a) Write a balanced chemical equation for this reaction, using molecular formulae.
   (b) Give the structural formula of chloroethane.
   (c) What type of reaction has taken place in this example?
5. Petrol (C_8H_{18}) is in fact not pure C_8H_{18} but a mixture of various *alkanes*. The 'octane rating' of petrol refers to the percentage of the petrol which is C_8H_{18}. For example, 93 octane fuel contains 93% C_8H_{18} and 7% other alkanes. The *isomer* of C_8H_{18} referred to in the 'octane rating' is in fact not octane but 2,2,4-trimethylpentane.
   (a) Write an unbalanced equation for the chemical reaction which takes place when petrol (C_8H_{18}) burns in excess oxygen.
   (b) Write the general formula of the *alkanes*.
   (c) Define the term *structural isomer*.
   (d) Use the information given in this question and your knowledge of naming organic compounds to deduce and draw the full structural formula for 2,2,4-trimethylpentane. (IEB pg 25)

9.7.5 The alkenes

In the alkenes, there is at least one double bond between two carbon atoms. This means that they are **unsaturated** and are *more reactive* than the alkanes. The simplest alkene is ethene (also known as ethylene), which is shown in figure 9.14.

(a)

(b) $\mathbf{CH_2CH_2}$ (c) $\mathbf{C_2H_4}$

Figure 9.14: The (a) structural, (b) condensed structural and (c) molecular structure representations of ethene

As with the alkanes, the elkenes also form a homologous series. They have the general formula C_nH_{2n}. The second alkene in the series would therefore be C_3H_6. This molecule is known as propene (figure 9.15). Note that if an alkene has two double bonds, it is called a **diene** and if it has three double bonds it is called a **triene**.

(a)

(b) $\mathbf{CH_3CHCH_2}$ (c) $\mathbf{C_3H_6}$

Figure 9.15: The (a) structural, (b) condensed structural and (c) molecular structure representations of propene

The elkenes have a variety of uses. Ethylene for example is a hormone in plants that stimulates the ripening of fruits and the opening of flowers. Propene is an important compound in the petrochemicals industry. It is used as a monomer to make polypropylene and is also used as a fuel gas for other industrial processes.

9.7.6 Naming the alkenes

Similar rules will apply in naming the alkenes, as for the alkanes.

Worked Example 42: Naming the alkenes

Question: Give the IUPAC name for the following compound:

Answer
Step 1 : Identify the functional group
The compound is an alkene and will have the suffix -ene.

Step 2 : Find the longest carbon chain
There are four carbon atoms in the longest chain and so the prefix for this compound will be 'but'.

Step 3 : Number the carbon atoms
Remember that when there is a double or triple bond, the carbon atoms must be numbered so that the double or triple bond is at the lowest numbered carbon. In this case, it doesn't matter whether we number the carbons from the left to right, or from the right to left. The double bond will still fall between C_2 and C_3. The position of the bond will come just before the suffix in the compound's name.

Step 4 : Look for any branched groups, name them and give their position on the carbon chain
There are no branched groups in this molecule.

Step 5 : Name the compound
The name of this compound is **but-2-ene**.

Worked Example 43: Naming the alkenes

Question: Draw the structural formula for the organic compound 3-methyl-butene

Answer
Step 1 : Identify the functional group
The suffix -ene means that this compound is an alkene and there must be a double bond in the molecule. There is no number immediately before the suffix which means that the double bond must be at the first carbon in the chain.

Step 2 : Determine the number of carbons in the longest chain
The prefix for the compound is 'but' so there must be four carbons in the longest chain.

Step 3 : Look for any branched groups
There is a methyl group at the third carbon atom in the chain.

Step 4 : Combine this information to draw the structural formula for this molecule.

Worked Example 44: Naming the alkenes

Question: Give the IUPAC name for the following compound:

Answer

Step 1 : Identify the functional group

The compound is an alkene and will have the suffix -ene. There is a double bond between the first and second carbons and also between the third and forth carbons. The organic compound is therefore a 'diene'.

Step 2 : Find the longest carbon chain and number the carbon atoms

There are four carbon atoms in the longest chain and so the prefix for this compound will be 'but'. The carbon atoms are numbered 1 to 4 in the diagram above. Remember that the main carbon chain must contain both the double bonds.

Step 3 : Look for any branched groups, name them and give their position on the carbon chain

There is a methyl group on the first carbon and an ethyl group on the second carbon.

Step 4 : Name the compound

The name of this compound is **1-methyl,2-ethyl-1,3 diene**.

Exercise: Naming the alkenes

Give the IUPAC name for each of the following alkenes:

1. C_5H_{10}
2. $CH_3CHCHCH_3$

3.

9.7.7 The properties of the alkenes

The properties of the alkenes are very similar to those of the alkanes, except that the alkenes are more reactive because they are unsaturated. As with the alkanes, compounds that have four or less carbon atoms are gases at room temperature, while those with five or more carbon atoms are liquids.

9.7.8 Reactions of the alkenes

Alkenes can undergo **addition reactions** because they are unsaturated. They readily react with hydrogen, water and the halogens. The double bond is broken and a single, saturated bond is formed. A new group is then added to one or both of the carbon atoms that previously made up the double bond. The following are some examples:

1. **Hydrogenation reactions**

   A catalyst such as platinum is normally needed for these reactions

   $CH_2 = CH_2 + H_2 \rightarrow CH_3 - CH_3$ (figure 9.16)

Figure 9.16: A hydrogenation reaction

2. **Halogenation reactions**

   $CH_2 = CH_2 + HBr \rightarrow CH_3 - CH_2 - Br$ (figure 9.17)

Figure 9.17: A halogenation reaction

3. **The formation of alcohols**

   $CH_2 = CH_2 + H_2O \rightarrow CH_3 - CH_2 - OH$ (figure 9.18)

Figure 9.18: The formation of an alcohol

Exercise: The Alkenes

1. Give the IUPAC name for each of the following organic compounds:

   (a)
   (b) CH_3CHCH_2

2. Refer to the data table below which shows the melting point and boiling point for a number of different organic compounds.

Formula	Name	Melting point (^{0}C)	Boiling point (^{0}C)
C_4H_{10}	Butane	-138	-0.5
C_5H_{12}	Pentane	-130	36
C_6H_{14}	Hexane	-95	69
C_4H_8	Butene	-185	-6
C_5H_{10}	Pentene	-138	30
C_6H_{12}	Hexene	-140	63

   (a) At room temperature (approx. 25^{0}C), which of the organic compounds in the table are:
      i. gases
      ii. liquids
   (b) In the alkanes...
      i. Describe what happens to the melting point and boiling point as the number of carbon atoms in the compound increases.
      ii. Explain why this is the case.
   (c) If you look at an alkane and an alkene that have the same number of carbon atoms...
      i. How do their melting points and boiling points compare?
      ii. Can you explain why their melting points and boiling points are different?
   (d) Which of the compounds, hexane or hexene, is more reactive? Explain your answer.

3. The following reaction takes place:

$$CH_3CHCH_2 + H_2 \rightarrow CH_3CH_2CH_3$$

   (a) Give the name of the organic compound in the reactants.
   (b) What is the name of the product?
   (c) What type of reaction is this?

(d) Which compound in the reaction is a saturated hydrocarbon?

9.7.9 The Alkynes

In the alkynes, there is at least one triple bond between two of the carbon atoms. They are unsaturated compounds and are therefore highly reactive. Their general formula is C_nH_{2n-2}. The simplest alkyne is ethyne (figure 9.19), also known as acetylene. Many of the alkynes are used to synthesise other chemical products.

Figure 9.19: Ethyne (acetylene)

Interesting Fact

The raw materials that are needed to make acetylene are calcium carbonate and coal. Acetylene can be produced through the following reactions:

$$CaCO_3 \rightarrow CaO$$
$$CaO + 3C \rightarrow CaC_2 + CO$$
$$CaC_2 + 2H_2O \rightarrow Ca(OH)_2 + C_2H_2$$

An important use of acetylene is in oxyacetylene gas welding. The fuel gas burns with oxygen in a torch. An incredibly high heat is produced, and this is enough to melt metal.

9.7.10 Naming the alkynes

The same rules will apply as for the alkanes and alkenes, except that the suffix of the name will now be -yne.

Worked Example 45: Naming the alkynes

Question: Give the IUPAC name for the following compound:

Answer

Step 1 : Identify the functional group

There is a triple bond between two of the carbon atoms, so this compound is an alkyne. The suffix will be -yne. The triple bond is at the second carbon, so the suffix will in fact be 2-yne.

Step 2 : Find the longest carbon chain and give the compound the correct prefix
If we count the carbons in a straight line, there are six. The prefix of the compound's name will be 'hex'.

Step 3 : Number the carbons in the longest chain
In this example, you will need to number the carbons from right to left so that the triple bond is between carbon atoms with the lowest numbers.

Step 4 : Look for any branched groups, name them and show the number of the carbon atom to which the group is attached
There is a methyl (CH_3) group attached to the fifth carbon (remember we have numbered the carbon atoms from right to left).

Step 5 : Combine the elements of the name into a single word in the following order: branched groups; prefix; name ending according to the functional group and its position along the longest carbon chain.
If we follow this order, the name of the compound is **5-methyl-hex-2-yne**.

Exercise: The alkynes
Give the IUPAC name for each of the following organic compounds.

1.
2. C_2H_2
3. CH_3CH_2CCH

9.8 The Alcohols

An alcohol is any organic compound where there is a *hydroxyl* functional group (-OH) bound to a carbon atom. The general formula for a simple alcohol is $C_nH_{2n+1}OH$.

The simplest and most commonly used alcohols are methanol and ethanol (figure 9.20).

The alcohols have a number of different uses:

- methylated spirits (surgical spirits) is a form of ethanol where methanol has been added
- ethanol is used in alcoholic drinks
- ethanol is used as an industrial solvent

Figure 9.20: (a) methanol and (b) ethanol

- methanol and ethanol can both be used as a fuel and they burn more cleanly than gasoline or diesel (refer to chapter 21 for more information on biofuels as an alternative energy resource.)
- ethanol is used as a solvent in medical drugs, perfumes and vegetable essences
- ethanol is an antiseptic

Interesting Fact

'Fermentation' refers to the conversion of sugar to alcohol using yeast (a fungus). The process of fermentation produces items such as wine, beer and yoghurt. To make wine, grape juice is fermented to produce alcohol. This reaction is shown below:

$$C_6H_{12}O_6 \rightarrow 2CO_2 + 2C_2H_5OH + energy$$

Interesting Fact

Ethanol is a diuretic. In humans, ethanol reduces the secretion of a hormone called antidiuretic hormone (ADH). The role of ADH is to control the amount of water that the body retains. When this hormone is not secreted in the right quantities, it can cause dehyration because too much water is lost from the body in the urine. This is why people who drink too much alcohol can become dehydrated, and experience symptoms such as headaches, dry mouth, and lethargy. Part of the reason for the headaches is that dehydration causes the brain to shrink away from the skull slightly. The effects of drinking too much alcohol can be reduced by drinking lots of water.

9.8.1 Naming the alcohols

The rules used to name the alcohols are similar to those already discussed for the other compounds, except that the suffix of the name will be different because the compound is an alcohol.

Worked Example 46: Naming alcohols 1

Question: Give the IUPAC name for the following organic compound
Answer
Step 1 : Identify the functional group
The compound has an -OH (hydroxyl) functional group and is therefore an alcohol. The compound will have the suffix -ol.

Step 2 : Find the longest carbon chain
There are three carbons in the longest chain. The prefix for this compound will be 'prop'. Since there are only single bonds between the carbon atoms, the suffix becomes 'propan' (similar to the alkane 'propane').

Step 3 : Number the carbons in the carbon chain
In this case, it doesn't matter whether you start numbering from the left or right. The hydroxyl group will still be attached to the middle carbon atom, numbered '2'.

Step 4 : Look for any branched groups, name them and give their position on the carbon chain.
There are no branched groups in this compound, but you still need to indicate the position of the hydroxyl group on the second carbon. The suffix will be -2-ol because the hydroxyl group is attached to the second carbon.

Step 5 : Combine the elements of the compound's name into a single word in the order of branched groups; prefix; name ending according to the functional group.
The compound's name is propan-2-ol.

Worked Example 47: Naming alcohols 2

Question: Give the IUPAC name for the following compound:

Answer
Step 1 : Identify the functional group
The compound has an -OH (hydroxyl) functional group and is therefore an alcohol. There are *two* hydroxyl groups in the compound, so the suffix will be -diol.

Step 2 : Find the longest carbon chain
There are four carbons in the longest chain. The prefix for this compound will be 'butan'.

Step 3 : Number the carbons in the carbon chain
The carbons will be numbered from left to right so that the two hydroxyl groups are attached to carbon atoms with the lowest numbers.

Step 4 : Look for any branched groups, name them and give their position on the carbon chain.
There are no branched groups in this compound, but you still need to indicate the position of the hydroxyl groups on the first and second carbon atoms. The suffix will therefore become 1,2-diol.

Step 5 : Combine the elements of the compound's name into a single word in the order of branched groups; prefix; name ending according to the functional group.
The compound's name is butan-1,2-diol.

Exercise: Naming the alcohols

1. Give the structural formula of each of the following organic compounds:
   (a) pentan-3-ol
   (b) butan-2,3-diol
   (c) 2-methyl-propanol
2. Give the IUPAC name for each of the following:
   (a) $CH_3CH_2CH(OH)CH_3$
   (b)

```
        H
        |
CH3— C —CH2—CH2—CH3
        |
        OH
```

9.8.2 Physical and chemical properties of the alcohols

The hydroxyl group affects the **solubility** of the alcohols. The hydroxyl group generally makes the alcohol molecule *polar* and therefore more likely to be **soluble** in water. However, the carbon chain resists solubility, so there are two opposing trends in the alcohols. Alcohols with shorter carbon chains are usually more soluble than those with longer carbon chains.

Alcohols tend to have higher **boiling points** than the hydrocarbons because of the strong hydrogen bond between hydrogen and oxygen atoms.

Alcohols can show either **acidic** or **basic** properties because of the hydroxyl group. They also undergo oxidation reactions to form aldehydes, ketones and carboxylic acids.

Activity :: Case Study : The uses of the alcohols

Read the extract below and then answer the questions that follow:

The alcohols are a very important group of organic compounds, and they have a variety of uses. Our most common use of the word 'alcohol' is with reference to alcoholic drinks. The alcohol in drinks is in fact **ethanol**. But ethanol has many more uses apart from alcoholic drinks! When ethanol burns in air, it produces carbon dioxide, water and energy and can therefore be used as a fuel on its own, or in mixtures with petrol. Because ethanol can be produced through fermentation, this is a useful way for countries without an oil industry to reduce imports of petrol. Ethanol is also used as a solvent in many perfumes and cosmetics.

Methanol can also be used as a fuel, or as a petrol additive to improve combustion. Most methanol is used as an industrial feedstock, in other words it is used to make other things such as methanal (formaldehyde), ethanoic acid and methyl esters. In most cases, these are then turned into other products.

Propan-2-ol is another important alcohol, which is used in a variety of applications as a solvent.

Questions

1. Give the structural formula for propan-2-ol.
2. Write a balanced chemical equation for the combustion reaction of ethanol.
3. Explain why the alcohols are good solvents.

9.9 Carboxylic Acids

Carboxylic acids are organic acids that are characterised by having a carboxyl group, which has the formula -(C=O)-OH, or more commonly written as -COOH. In a carboxyl group, an oxygen atom is double-bonded to a carbon atom, which is also bonded to a hydroxyl group. The simplest carboxylic acid, methanoic acid, is shown in figure 9.21.

Figure 9.21: Methanoic acid

Carboxylic acids are widespread in nature. Methanoic acid (also known as *formic acid*) has the formula HCOOH and is found in insect stings. Ethanoic acid (CH_3COOH), or *acetic acid*, is the main component of vinegar. More complex organic acids also have a variety of different functions. Benzoic acid (C_6H_5COOH) for example, is used as a food preservative.

A certain type of ant, called formicine ants, manufacture and secrete formic acid, which is used to defend themselves against other organisms that might try to eat them.

9.9.1 Physical Properties

Carboxylic acids are **weak acids**, in other words they only dissociate partially. Why does the carboxyl group have acidic properties? In the carboxyl group, the hydrogen tends to separate itself from the oxygen atom. In other words, the carboxyl group becomes a source of positively-charged hydrogen ions (H^+). This is shown in figure 9.22.

Figure 9.22: The dissociation of a carboxylic acid

Exercise: Carboxylic acids

1. Refer to the table below which gives information about a number of carboxylic acids, and then answer the questions that follow.

Formula	Common name	Source	IUPAC name	melting point (^{0}C)	boiling point (^{0}C)
	formic acid	ants	methanoic acid	8.4	101
CH_3CO_2H		vinegar	ethanoic acid	16.6	118
	propionic acid	milk	propanoic acid	-20.8	141
$CH_3(CH_2)_2CO_2H$	butyric acid	butter		-5.5	164
	valeric acid	valerian root	pentanoic acid	-34.5	186
$CH_3(CH_2)_4CO_2H$	caproic acid	goats		-4	205
	enanthic acid	vines		-7.5	223
$CH_3(CH_2)_6CO_2H$	caprylic acid	goats		16.3	239

(a) Fill in the missing spaces in the table by writing the formula, common name or IUPAC name.

(b) Draw the structural formula for butyric acid.

(c) Give the molecular formula for caprylic acid.

(d) Draw a graph to show the relationship between molecular mass (on the x-axis) and boiling point (on the y-axis)

i. Describe the trend you see.
ii. Suggest a reason for this trend.

9.9.2 Derivatives of carboxylic acids: The esters

When an alcohol reacts with a carboxylic acid, an **ester** is formed. Most esters have a characteristic and pleasant smell. In the reaction, the hydrogen atom from the hydroxyl group, and an OH from the carboxlic acid, form a molecule of water. A new bond is formed between what remains of the alcohol and acid. The name of the ester is a combination of the names of the alcohol and carboxylic acid. The suffix for an ester is -oate. An example is shown in figure 9.23.

Figure 9.23: The formation of an ester from an alcohol and carboxylic acid

9.10 The Amino Group

The amino group has the formula -NH_2 and consists of a nitrogen atom that is bonded to two hydrogen atoms, and to the carbon skeleton. Organic compounds that contain this functional group are called **amines**. One example is *glycine*. Glycine belongs to a group of organic compounds called *amino acids*, which are the building blocks of proteins.

Figure 9.24: A molecule of glycine

9.11 The Carbonyl Group

The carbonyl group (-CO) consists of a carbon atom that is joined to an oxygen by a double bond. If the functional group is on the *end* of the carbon chain, the organic compound is called a **ketone**. The simplest ketone is *acetone*, which contains three carbon atoms. A ketone has the ending 'one' in its IUPAC name.

Exercise: Carboxylic acids, esters, amines and ketones

1. Look at the list of organic compounds in the table below:

Organic compound	Type of compound
$CH_3CH_2CH_2COOH$	
NH_2CH_2COOH	
propyl ethanoate	
CH_3CHO	

   (a) Complete the table by identifying each compound as either a carboxylic acid, ester, amine or ketone.
   (b) Give the name of the compounds that have been written as condensed structural formulae.

2. A chemical reaction takes place and ethyl methanoate is formed.
   (a) What type of organic compound is ethyl methanoate?
   (b) Name the two reactants in this chemical reaction.
   (c) Give the structural formula of ethyl methanoate.

9.12 Summary

- **Organic chemistry** is the branch of chemistry that deals with organic molecules. An **organic molecule** is one that contains carbon.
- All **living organisms** contain carbon. Plants use sunlight to convert carbon dioxide in the air into organic compounds through the process of **photosynthesis**. Animals and other organisms then feed on plants to obtain their own organic compounds. **Fossil fuels** are another important source of carbon.
- It is the **unique properties of the carbon atom** that give organic compounds certain properties.
- The carbon atom has **four valence electrons**, so it can bond with many other atoms, often resulting in long chain structures. It also forms mostly **covalent bonds** with the atoms that it bonds to, meaning that most organic molecules are **non-polar**.
- An organic compound can be represented in different ways, using its **molecular formula**, **structural formula** or **condensed structural formula**.
- If two compounds are **isomers**, it means that they have the same molecular formulae but different structural formulae.
- A **functional group** is a particular group of atoms within a molecule, which give it certain reaction characteristics. Organic compounds can be grouped according to their functional group.
- The **hydrocarbons** are organic compounds that contain only carbon and hydrogen. They can be further divided into the alkanes, alkenes and alkynes, based on the type of bonds between the carbon atoms.
- The **alkanes** have only **single bonds** between their carbon atoms and are unreactive.
- The **alkenes** have at least one **double bond** between two of their carbon atoms. They are more reactive than the alkanes.
- The **alkynes** have at least one **triple bond** between two of their carbon atoms. They are the most reactive of the three groups.
- A hydrocarbon is said to be **saturated** if it contains the maximum possible number of hydrogen atoms for that molecule. The alkanes are all saturated compounds.

- A hydrocarbon is **unsaturated** if it does not contain the maximum number of hydrogen atoms for that molecule. The alkenes and alkynes are examples of unsaturated molecules. If a double or triple bond is broken, more hydrogen atoms can be added to the molecule.
- There are three types of reactions that occur in the alkanes: **substitution**, **elimination** and **oxidation** reactions.
- The alkenes undergo **addition** reactions because they are unsaturated.
- Organic compounds are **named** according to their functional group and its position in the molecule, the number of carbon atoms in the molecule and the position of any double and triple bonds. The IUPAC rules for nomenclature are used in the naming of organic molecules.
- Many of the **properties** of the hydrocarbons are determined by their **molecular structure**, the **bonds** between atoms and molecules, and their **surface area**.
- The **melting point** and **boiling point** of the hydrocarbons increases as their number of carbon atoms increases.
- The **molecular mass** of the hydrocarbons determines whether they will be in the gaseous, liquid or solid phase at certain temperatures.
- An **alcohol** is an organic compound that contains a **hydroxyl group** (OH).
- The alcohols have a number of different uses including their use as a solvent, for medicinal purposes and in alcoholic drinks.
- The alcohols share a number of **properties** because of the hydroxyl group. The hydroxyl group affects the **solubility** of the alcohols. Those with shorter carbon chains are generally more soluble, and those with longer chains are less soluble. The strong hydrogen bond between the hydrogen and oxygen atoms in the hydroxyl group gives alcohols a higher melting point and boiling point than other organic compounds. The hydroxyl group also gives the alcohols both acidic and basic properties.
- The **carboxylic acids** are organic acids that contain a **carboxyl group** with the formula COOH. In a carboxyl group, an oxygen atom is double-bonded to a carbon atom, which is also bonded to a hydroxyl group.
- The carboxylic acids have weak **acidic properties** because the hydrogen atom is able to dissociate from the carboxyl group.
- An **ester** is formed when an alcohol reacts with a carboxylic acid.
- The **amines** are organic compounds that contain an **amino** functional group, which has the formula NH_2. Some amines belong to the **amino acid** group, which are the building blocks of proteins.
- The **ketones** are a group of compounds that contain a **carbonyl group**, which consists of an oxygen atom that is double-bonded to a carbon atom. In a ketone, the carbonyl group is on the end of the carbon chain.

Exercise: Summary exercise

1. Give **one word** for each of the following descriptions:
   (a) The group of hydrocarbons to which 2-methyl-propene belongs.
   (b) The name of the functional group that gives alcohols their properties.
   (c) The group of organic compounds that have acidic properties.
   (d) The name of the organic compound that is found in vinegar.
   (e) The name of the organic compound that is found in alcoholic beverages.

2. In each of the following questions, choose the **one correct answer** from the list provided.

   (a) When 1-propanol is oxidised by acidified potassium permanganate, the possible product formed is...

   i. propane
   ii. propanoic acid
   iii. methyl propanol
   iv. propyl methanoate

   (IEB 2004)

   (b) What is the IUPAC name for the compound represented by the following structural formula?

   i. 1,2,2-trichlorobutane
   ii. 1-chloro-2,2-dichlorobutane
   iii. 1,2,2-trichloro-3-methylpropane
   iv. 1-chloro-2,2-dichloro-3-methylpropane

   (IEB 2003)

3. Write balanced equations for the following reactions:

   (a) Ethene reacts with bromine
   (b) Ethyne gas burns in an excess of oxygen
   (c) Ethanoic acid ionises in water

4. The table below gives the boiling point of ten organic compounds.

	Compound	**Formula**	**Boiling Point (^{0}C)**
1	methane	CH_4	-164
2	ethane	C_2H_6	-88
3	propane	C_3H_8	-42
4	butane	C_4H_{10}	0
5	pentane	C_5H_{12}	36
6	methanol	CH_3OH	65
7	ethanol	C_2H_5OH	78
8	propan-1-ol	C_3H_7OH	98
9	propan-1,2-diol	$CH_3CHOHCH_2OH$	189
10	propan-1,2,3-triol	$CH_2OHCHOHCH_2OH$	290

   The following questions refer to the compounds shown in the above table.

   (a) To which homologous series do the following compounds belong?

   i. Compounds 1,2 and 3
   ii. Compounds 6,7 and 8

   (b) Which of the above compounds are gases at room temperature?
   (c) What causes the trend of increasing boiling points of compounds 1 to 5?
   (d) Despite the fact that the length of the carbon chain in compounds 8,9 and 10 is the same, the boiling point of propan-1,2,3-triol is much higher than the boiling point of propan-1-ol. What is responsible for this large difference in boiling point?
   (e) Give the IUPAC name and the structural formula of an isomer of butane.

(f) Which **one** of the above substances is used as a reactant in the preparation of the ester ethylmethanoate?

(g) Using structural formulae, write an equation for the reaction which produces ethylmethanoate.

(*IEB 2004*)

5. Refer to the numbered diagrams below and then answer the questions that follow.

(a) Which one of the above compounds is produced from the fermentation of starches and sugars in plant matter?
   i. compound 1
   ii. compound 2
   iii. compound 3
   iv. compound 4

(b) To which one of the following homologous series does compound 1 belong?
   i. esters
   ii. alcohols
   iii. aldehydes
   iv. carboxylic acids

(c) The correct IUPAC name for compound 3 is...
   i. 1,1-dibromo-3-butyne
   ii. 4,4-dibromo-1-butyne
   iii. 2,4-dibromo-1-butyne
   iv. 4,4-dibromo-1-propyne

(d) What is the correct IUPAC name for compound 4?
   i. propanoic acid
   ii. ethylmethanoate
   iii. methylethanoate
   iv. methylpropanoate

IEB 2005

6. Answer the following questions:

(a) What is a homologous series?

(b) A mixture of ethanoic acid and methanol is warmed in the presence of concentrated sulphuric acid.
   i. Using structural formulae, give an equation for the reaction which takes place.

ii. What is the IUPAC name of the organic compound formed in this reaction?

(c) Consider the following unsaturated hydrocarbon:

i. Give the IUPAC name for this compound.

ii. Give the balanced equation for the combustion of this compound in excess oxygen.

(IEB Paper 2, 2003)

7. Consider the organic compounds labelled A to E.
A. $CH_3CH_2CH_2CH_2CH_2CH_3$
B. C_6H_6
C. CH_3-Cl
D. Methylamine
E

(a) Write a balanced chemical equation for the preparation of compound C using an alkane as one of the reactants.

(b) Write down the IUPAC name for compound E.

(c) Write down the structural formula of an isomer of compound A that has only FOUR carbon atoms in the longest chain.

(d) Write down the structural formula for compound B.

Chapter 10

Organic Macromolecules - Grade 12

As its name suggests, a macromolecule is a large molecule that forms when lots of smaller molecules are joined together. In this chapter, we will be taking a closer look at the structure and properties of different macromolecules, and at how they form.

10.1 Polymers

Some macromolecules are made up of lots of repeating structural units called **monomers**. To put it more simply, a monomer is like a building block. When lots of similar monomers are joined together by covalent bonds, they form a **polymer**. In an **organic polymer**, the monomers would be joined by the *carbon* atoms of the polymer 'backbone'. A polymer can also be **inorganic**, in which case there may be atoms such as *silicon* in the place of carbon atoms. The key feature that makes a polymer different from other macromolecules, is the repetition of identical or similar monomers in the polymer chain. The examples shown below will help to make these concepts clearer.

Definition: Polymer
Polymer is a term used to describe large molecules consisting of repeating structural units, or monomers, connected by covalent chemical bonds.

1. **Polyethene**

   Chapter 9 looked at the structure of a group of hydrocarbons called the *alkenes*. One example is the molecule **ethene**. The structural formula of ethene is is shown in figure 10.1. When lots of ethene molecules bond together, a polymer called **polyethene** is formed. Ethene is the *monomer* which, when joined to other ethene molecules, forms the *polymer* **polyethene**. Polyethene is a cheap plastic that is used to make plastic bags and bottles.

Figure 10.1: (a) Ethene monomer and (b) polyethene polymer

A polymer may be a chain of thousands of monomers, and so it is impossible to draw the entire polymer. Rather, the structure of a polymer can be condensed and represented as shown in figure 10.2. The monomer is enclosed in brackets and the 'n' represents the number of ethene molecules in the polymer, where 'n' is any whole number. What this shows is that the ethene monomer is repeated an indefinite number of times in a molecule of polyethene.

Figure 10.2: A simplified representation of a polyethene molecule

2. **Polypropene**

   Another example of a polymer is *polypropene* (fig 10.3). Polypropene is also a plastic, but is stronger than polyethene and is used to make crates, fibres and ropes. In this polymer, the monomer is the alkene called **propene**.

Figure 10.3: (a) Propene monomer and (b) polypropene polymer

10.2 How do polymers form?

Polymers are formed through a process called **polymerisation**, where monomer molecules react together to form a polymer chain. Two types of polymerisation reactions are **addition polymerisation** and **condensation polymerisation**.

Definition: Polymerisation
In chemistry, polymerisation is a process of bonding monomers, or *single units* together through a variety of reaction mechanisms to form longer chains called polymers.

10.2.1 Addition polymerisation

In this type of reaction, monomer molecules are added to a growing polymer chain one at a time. No small molecules are eliminated in the process. An example of this type of reaction is the formation of *polyethene* from *ethene* (fig 10.1). When molecules of ethene are joined to each other, the only thing that changes is that the double bond between the carbon atoms in each ethene monomer is replaced by a single bond so that a new carbon-carbon bond can be formed with the next monomer in the chain. In other words, the monomer is an *unsaturated* compound which, after an addition reaction, becomes a *saturated* compound.

Extension: Initiation, propagation and termination

There are three stages in the process of addition polymerisation. **Initiation** refers to a chemical reaction that triggers off another reaction. In other words, initiation is the starting point of the polymerisation reaction. **Chain propagation** is the part where monomers are continually added to form a longer and longer polymer chain. During chain propagation, it is the reactive end groups of the polymer chain that react in each propagation step, to add a new monomer to the chain. Once a monomer has been added, the reactive part of the polymer is now in this last monomer unit so that propagation will continue. **Termination** refers to a chemical reaction that destroys the reactive part of the polymer chain so that propagation stops.

Worked Example 48: Polymerisation reactions

Question: A polymerisation reaction takes place and the following polymer is formed:

Note: W, X, Y and Z could represent a number of different atoms or combinations of atoms e.g. H, F, Cl or CH_3.

1. Give the structural formula of the monomer of this polymer.
2. To what group of organic compounds does this monomer belong?
3. What type of polymerisation reaction has taken place to join these monomers to form the polymer?

Answer

Step 1 : Look at the structure of the repeating unit in the polymer to determine the monomer.

The monomer is:

Step 2 : Look at the atoms and bonds in the monomer to determine which group of organic compounds it belongs to.

The monomer has a double bond between two carbon atoms. The monomer must be an alkene.

Step 3 : Determine the type of polymerisation reaction.

In this example, unsaturated monomers combine to form a saturated polymer. No atoms are lost or gained for the bonds between monomers to form. They are simply added to each other. This is an addition reaction.

10.2.2 Condensation polymerisation

In this type of reaction, two monomer molecules form a covalent bond and a small molecule such as water is lost in the bonding process. Nearly all biological reactions are of this type. **Polyester** and **nylon** are examples of polymers that form through condensation polymerisation.

1. **Polyester**

   Polyesters are a group of polymers that contain the **ester** functional group in their main chain. Although there are many forms of polyesters, the term *polyester* usually refers to polyethylene terephthalate (PET). PET is made from ethylene glycol (an alcohol) and terephthalic acid (an acid). In the reaction, a hydrogen atom is lost from the alcohol, and a hydroxyl group is lost from the carboxylic acid. Together these form one water molecule which is lost during condensation reactions. A new bond is formed between an oxygen and a carbon atom. This bond is called an **ester linkage**. The reaction is shown in figure 10.4.

Figure 10.4: An acid and an alcohol monomer react (a) to form a molecule of the polyester 'polyethylene terephthalate' (b).

Polyesters have a number of characteristics which make them very useful. They are resistant to stretching and shrinking, they are easily washed and dry quickly, and they are resistant to mildew. It is for these reasons that polyesters are being used more and more in **textiles**. Polyesters are stretched out into fibres and can then be made into fabric and articles of clothing. In the home, polyesters are used to make clothing, carpets, curtains, sheets, pillows and upholstery.

Interesting Fact

Polyester is not just a textile. Polyethylene terephthalate is in fact a plastic which can also be used to make plastic drink bottles. Many drink bottles are recycled by being reheated and turned into polyester fibres. This type of recycling helps to reduce disposal problems.

2. **Nylon**

   Nylon was the first polymer to be commercially successful. Nylon replaced silk, and was used to make parachutes during World War 2. Nylon is very strong and resistant, and is used in fishing line, shoes, toothbrush bristles, guitar strings and machine parts to name

just a few. Nylon is formed from the reaction of an amine (1,6-diaminohexane) and an acid monomer (adipic acid) (figure 10.5). The bond that forms between the two monomers is called an **amide linkage**. An amide linkage forms between a nitrogen atom in the amine monomer and the carbonyl group in the carboxylic acid.

Figure 10.5: An amine and an acid monomer (a) combine to form a section of a nylon polymer (b).

Interesting Fact

Nylon was first introduced around 1939 and was in high demand to make stockings. However, as World War 2 progressed, nylon was used more and more to make parachutes, and so stockings became more difficult to buy. After the war, when manufacturers were able to shift their focus from parachutes back to stockings, a number of riots took place as women queued to get stockings. In one of the worst disturbances, 40 000 women queued up for 13 000 pairs of stockings, which led to fights breaking out!

Exercise: Polymers

1. The following monomer is a reactant in a polymerisation reaction:

   (a) What is the IUPAC name of this monomer?
   (b) Give the structural formula of the polymer that is formed in this polymerisation reaction.
   (c) Is the reaction an addition or condensation reaction?

2. The polymer below is the product of a polymerisation reaction.

   (a) Give the structural formula of the monomer in this polymer.
   (b) What is the name of the monomer?
   (c) Draw the abbreviated structural formula for the polymer.
   (d) Has this polymer been formed through an addition or condensation polymerisation reaction?

3. A condensation reaction takes place between methanol and methanoic acid.
   (a) Give the structural formula for...
      i. methanol
      ii. methanoic acid
      iii. the product of the reaction
   (b) What is the name of the product? (Hint: The product is an ester)

10.3 The chemical properties of polymers

The attractive forces between polymer chains play a large part in determining a polymer's properties. Because polymer chains are so long, these interchain forces are very important. It is usually the side groups on the polymer that determine what types of intermolecular forces will exist. The greater the strength of the intermolecular forces, the greater will be the tensile strength and melting point of the polymer. Below are some examples:

- **Hydrogen bonds between adjacent chains**

  Polymers that contain amide or carbonyl groups can form *hydrogen bonds* between adjacent chains. The positive hydrogen atoms in the N-H groups of one chain are strongly attracted to the oxygen atoms in the C=O groups on another. Polymers that contain *urea* linkages would fall into this category. The structural formula for urea is shown in figure 10.6. Polymers that contain urea linkages have high tensile strength and a high melting point.

Figure 10.6: The structural formula for urea

- **Dipole-dipole bonds between adjacent chains**

  Polyesters have *dipole-dipole bonding* between their polymer chains. Dipole bonding is not as strong as hydrogen bonding, so a polyester's melting point and strength are lower

than those of the polymers where there are hydrogen bonds between the chains. However, the weaker bonds between the chains means that polyesters have greater flexibility. The greater the flexibility of a polymer, the more likely it is to be moulded or stretched into fibres.

- **Weak van der Waal's forces**

  Other molecules such as ethene do not have a permanent dipole and so the attractive forces between polyethene chains arise from weak *van der Waals* forces. Polyethene therefore has a lower melting point than many other polymers.

10.4 Types of polymers

There are many different types of polymers. Some are organic, while others are inorganic. Organic polymers can be broadly grouped into either synthetic/semi-synthetic (artificial) or biological (natural) polymers. We are going to take a look at two groups of organic polymers: *plastics*, which are usually synthetic or semi-synthetic and *biological macromolecules* which are natural polymers. Both of these groups of polymers play a very important role in our lives.

10.5 Plastics

In today's world, we can hardly imagine life without plastic. From cellphones to food packaging, fishing line to plumbing pipes, compact discs to electronic equipment, plastics have become a very important part of our daily lives. "Plastics" cover a range of synthetic and semi-synthetic organic polymers. Their name comes from the fact that they are 'malleable', in other words their shape can be changed and moulded.

Definition: Plastic
Plastic covers a range of synthetic or semisynthetic organic polymers. Plastics may contain other substances to improve their performance. Their name comes from the fact that many of them are malleable, in other words they have the property of plasticity.

It was only in the nineteenth century that it was discovered that plastics could be made by chemically changing natural polymers. For centuries before this, only natural organic polymers had been used. Examples of natural organic polymers include *waxes* from plants, *cellulose* (a plant polymer used in fibres and ropes) and *natural rubber* from rubber trees. But in many cases, these natural organic polymers didn't have the characteristics that were needed for them to be used in specific ways. Natural rubber for example, is sensitive to temperature and becomes sticky and smelly in hot weather and brittle in cold weather.

In 1834 two inventors, Friedrich Ludersdorf of Germany and Nathaniel Hayward of the US, independently discovered that adding sulfur to raw rubber helped to stop the material from becoming sticky. After this, Charles Goodyear discovered that heating this modified rubber made it more resistant to abrasion, more elastic and much less sensitive to temperature. What these inventors had done was to improve the properties of a natural polymer so that it could be used in new ways. An important use of rubber now is in vehicle tyres, where these properties of rubber are critically important.

Interesting Fact

The first true plastic (i.e. one that was not based on any material found in nature) was *Bakelite*, a cheap, strong and durable plastic. Some of these plastics are still used for example in electronic circuit boards, where their properties of insulation and heat resistance are very important.

10.5.1 The uses of plastics

There is such a variety of different plastics available, each having their own specific properties and uses. The following are just a few examples.

- **Polystyrene**

  Polystyrene (figure 15.2) is a common plastic that is used in model kits, disposable eating utensils and a variety of other products. In the polystyrene polymer, the monomer is *styrene*, a liquid hydrocarbon that is manufactured from petroleum.

Figure 10.7: The polymerisation of a styrene monomer to form a polystyrene polymer

- **Polyvinylchloride (PVC)**

  Polyvinyl chloride (PVC) (figure 10.8) is used in plumbing, gutters, electronic equipment, wires and food packaging. The side chains of PVC contain chlorine atoms, which give it its particular characteristics.

Figure 10.8: Polyvinyl chloride

Interesting Fact

Many vinyl products have other chemicals added to them to give them particular properties. Some of these chemicals, called additives, can leach out of the vinyl products. In PVC, *plasticizers* are used to make PVC more flexible. Because many baby toys are made from PVC, there is concern that some of these products may leach into the mouths of the babies that are chewing on them. In the USA, most companies have stopped making PVC toys. There are also concerns that some of the plasticizers added to PVC may cause a number of health conditions including cancer.

- **Synthetic rubber**

  Another plastic that was critical to the World War 2 effort was *synthetic rubber*, which was produced in a variety of forms. Not only were worldwide natural rubber supplies limited,

but most rubber-producing areas were under Japanese control. Rubber was needed for tyres and parts of war machinery. After the war, synthetic rubber also played an important part in the space race and nuclear arms race.

- **Polyethene/polyethylene (PE)**

  Polyethylene (figure 10.1) was discovered in 1933. It is a cheap, flexible and durable plastic and is used to make films and packaging materials, containers and car fittings. One of the most well known polyethylene products is 'Tupperware', the sealable food containers designed by Earl Tupper and promoted through a network of housewives!

- **Polytetrafluoroethylene (PTFE)**

  Polytetrafluoroethylene (figure 10.9) is more commonly known as 'Teflon' and is most well known for its use in non-stick frying pans. Teflon is also used to make the breathable fabric Gore-Tex.

Figure 10.9: A tetra fluoroethylene monomer and polytetrafluoroethylene polymer

Table 10.1 summarises the formulae, properties and uses of some of the most common plastics.

Table 10.1: A summary of the formulae, properties and uses of some common plastics

Name	Formula	Monomer	Properties	Uses
Polyethene (low density)	-(CH_2-CH_2)$_n$-	CH_2=CH_2	soft, waxy solid	film wrap and plastic bags
Polyethene (high density)	-(CH_2-CH_2)$_n$-	CH_2=CH_2	rigid	electrical insulation, bottles and toys
Polypropene	-[CH_2-CH(CH_3)]$_n$-	CH_2=$CHCH_3$	different grades: some are soft and others hard	carpets and upholstery
Polyvinylchloride (PVC)	-(CH_2-CHCl)$_n$-	CH_2=CHCl	strong, rigid	pipes, flooring
Polystyrene	-[CH_2-CH(C_6H_5)]$_n$	CH_2=CHC_6H_5	hard, rigid	toys, packaging
Polytetrafluoroethylene	-(CF_2-CF_2)$_n$-	CF_2=CF_2	resistant, smooth, solid	non-stick surfaces, electrical insulation

---

?

Exercise: Plastics

1. It is possible for macromolecules to be composed of more than one type of repeating monomer. The resulting polymer is called a **copolymer**. Varying the monomers that combine to form a polymer, is one way of controlling the properties of the resulting material. Refer to the table below which shows a number of different copolymers of rubber, and answer the questions that follow:

Monomer A	Monomer B	Copolymer	Uses
$H_2C{=}CHCl$	$H_2C{=}CCl_2$	Saran	films and fibres
$H_2C{=}CHC_6H_5$	$H_2C{=}C\text{-}CH{=}CH_2$	SBR (styrene butadiene rubber)	tyres
$H_2C{=}CHCN$	$H_2C{=}C\text{-}CH{=}CH_2$	Nitrile rubber	adhesives and hoses
$H_2C{=}C(CH_3)_2$	$H_2C{=}C\text{-}CH{=}CH_2$	Butyl rubber	inner tubes
$F_2C{=}CF(CF_3)$	$H_2C{=}CHF$	Viton	gaskets

(a) Give the structural formula for each of the monomers of nitrile rubber.
(b) Give the structural formula of the copolymer viton.
(c) In what ways would you expect the properties of SBR to be different from nitrile rubber?
(d) Suggest a reason why the properties of these polymers are different.

2. In your home, find as many examples of different types of plastics that you can. Bring them to school and show them to your group. Together, use your examples to complete the following table:

Object	Type of plastic	Properties	Uses

10.5.2 Thermoplastics and thermosetting plastics

A **thermoplastic** is a plastic that can be melted to a liquid when it is heated and freezes to a brittle, glassy state when it is cooled enough. These properties of thermoplastics are mostly due to the fact that the forces between chains are weak. This also means that these plastics can be easily stretched or moulded into any shape. Examples of thermoplastics include nylon, polystyrene, polyethylene, polypropylene and PVC. Thermoplastics are more easily recyclable than some other plastics.

Thermosetting plastics differ from thermoplastics because once they have been formed, they cannot be remelted or remoulded. Examples include bakelite, vulcanised rubber, melanine (used to make furniture), and many glues. Thermosetting plastics are generally stronger than thermoplastics and are better suited to being used in situations where there are high temperatures. They are not able to be recycled. Thermosetting plastics have strong covalent bonds between chains and this makes them very strong.

Activity :: Case Study : Biodegradable plastics

Read the article below and then answer the questions that follow.

Our whole world seems to be wrapped in plastic. Almost every product we buy, most of the food we eat and many of the liquids we drink come encased in plastic. Plastic packaging provides excellent protection for the product, it is cheap to manufacture and seems to last forever. Lasting forever, however, is proving to be a major environmental problem. Another problem is that traditional plastics are manufactured from non-renewable resources - oil, coal and natural gas. In an effort to overcome these problems, researchers and engineers have been trying to develop biodegradable plastics that are made from renewable resources, such as plants.

The term biodegradable means that a substance can be broken down into simpler substances by the activities of living organisms, and therefore is unlikely to remain in the environment. The reason most plastics are not biodegradable is because their long polymer molecules are too large and too tightly bonded together to be broken apart and used by decomposer organisms. However, plastics based on natural plant polymers that come from wheat or corn starch have molecules that can be more easily broken down by microbes.

Starch is a natural polymer. It is a white, granular carbohydrate produced by plants during photosynthesis and it serves as the plant's energy store. Many plants contain large amounts of starch. Starch can be processed directly into a bioplastic but, because it is soluble in water, articles made from starch will swell and deform when exposed to moisture, and this limits its use. This problem can be overcome by changing starch into a different polymer. First, starch is harvested from corn, wheat or potatoes, then microorganisms transform it into lactic acid, a monomer. Finally, the lactic acid is chemically treated to cause the molecules of lactic acid to link up into long chains or polymers, which bond together to form a plastic called polylactide (PLA).

PLA can be used for products such as plant pots and disposable nappies. It has been commercially available in some countries since 1990, and certain blends have proved successful in medical implants, sutures and drug delivery systems because they are able to dissolve away over time. However, because PLA is much more expensive than normal plastics, it has not become as popular as one would have hoped.

Questions

1. In your own words, explain what is meant by a 'biodegradable plastic'.
2. Using your knowledge of chemical bonding, explain why some polymers are biodegradable and others are not.
3. Explain why lactic acid is a more useful monomer than starch, when making a biodegradable plastic.
4. If you were a consumer (shopper), would you choose to buy a biodegradable plastic rather than another? Explain your answer.
5. What do you think could be done to make biodegradable plastics more popular with consumers?

10.5.3 Plastics and the environment

Although plastics have had a huge impact globally, there is also an environmental price that has to be paid for their use. The following are just some of the ways in which plastics can cause damage to the environment.

1. **Waste disposal**

   Plastics are not easily broken down by micro-organisms and therefore most are not easily biodegradeable. This leads to waste dispoal problems.

2. **Air pollution**

   When plastics burn, they can produce toxic gases such as carbon monoxide, hydrogen cyanide and hydrogen chloride (particularly from PVC and other plastics that contain chlorine and nitrogen).

3. **Recycling**

   It is very difficult to recycle plastics because each type of plastic has different properties and so different recycling methods may be needed for each plastic. However, attempts are being made to find ways of recycling plastics more effectively. Some plastics can be remelted and re-used, while others can be ground up and used as a filler. However,

one of the problems with recycling plastics is that they have to be sorted according to plastic *type*. This process is difficult to do with machinery, and therefore needs a lot of labour. Alternatively, plastics should be re-used. In many countries, including South Africa, shoppers must now pay for plastic bags. This encourages people to collect and re-use the bags they already have.

Activity :: Case Study : Plastic pollution in South Africa

Read the following extract, taken from 'Planet Ark' (September 2003), and then answer the questions that follow.

> South Africa launches a programme this week to exterminate its "national flower" - the millions of used plastic bags that litter the landscape.
> Beginning on Friday, plastic shopping bags used in the country must be both thicker and more recyclable, a move officials hope will stop people from simply tossing them away. "Government has targeted plastic bags because they are the most visible kind of waste," said Phindile Makwakwa, spokeswoman for the Department of Environmental Affairs and Tourism. "But this is mostly about changing people's mindsets about the environment."
> South Africa is awash in plastic pollution. Plastic bags are such a common eyesore that they are dubbed "roadside daisies" and referred to as the national flower. Bill Naude of the Plastics Federation of South Africa said the country used about eight billion plastic bags annually, a figure which could drop by 50 percent if the new law works.

It is difficult sometimes to imagine exactly how much waste is produced in our country every year. Where does all of this go to? You are going to do some simple calculations to try to estimate the volume of plastic packets that is produced in South Africa every year.

1. Take a plastic shopping packet and squash it into a tight ball.
   (a) Measure the approximate length, breadth and depth of your squashed plastic bag.
   (b) Calculate the approximate volume that is occupied by the packet.
   (c) Now calculate the approximate volume of your classroom by measuring its length, breadth and height.
   (d) Calculate the number of squashed plastic packets that would fit into a classroom of this volume.
   (e) If South Africa produces an average of 8 billion plastic bags each year, how many clasrooms would be filled if all of these bags were thrown away and not re-used?
2. What has South Africa done to try to reduce the number of plastic bags that are produced?
3. Do you think this has helped the situation?
4. What can *you* do to reduce the amount of plastic that you throw away?

10.6 Biological Macromolecules

A *biological macromolecule* is one that is found in living organisms. Biological macromolecules include molecules such as carbohydrates, proteins and nucleic acids. Lipids are also biological macromolecules. They are essential for all forms of life to survive.

Definition: Biological macromolecule
A biological macromolecule is a polymer that occurs naturally in living organisms. These molecules are essential to the survival of life.

10.6.1 Carbohydrates

Carbohydrates include the sugars and their polymers. One key characteristic of the carbohydrates is that they contain only the elements carbon, hydrogen and oxygen. In the carbohydrate monomers, every carbon except one has a hydroxyl group attached to it, and the remaining carbon atom is double bonded to an oxygen atom to form a carbonyl group. One of the most important monomers in the carbohydrates is **glucose** (figure 10.10). The glucose molecule can exist in an open-chain (acyclic) and ring (cyclic) form.

Figure 10.10: The open chain (a) and cyclic (b) structure of a glucose molecule

Glucose is produced during **photosynthesis**, which takes place in plants. During photosynthesis, sunlight (solar energy), water and carbon dioxide are involved in a chemical reaction that produces glucose and oxygen. This glucose is stored in various ways in the plant.

The photosynthesis reaction is as follows:

$$6CO_2 + 6H_2O + \text{sunlight} \rightarrow C_6H_{12}O_6 + 6O_2$$

Glucose is an important source of **energy** for both the plant itself, and also for the other animals and organisms that may feed on it. Glucose plays a critical role in **cellular respiration**, which is a chemical reaction that occurs in the cells of all living organisms. During this reaction, glucose and oxygen react to produce carbon dioxide, water and ATP energy. ATP is a type of energy that can be used by the body's cells so that they can function normally. The purpose of *eating* then, is to obtain glucose which the body can then convert into the ATP energy it needs to be able to survive.

The reaction for cellular respiration is as follows:

$$6C_6H_{12}O_6 + 60_2 \rightarrow 6CO_2 + 6H_2O + \text{ATP (cell energy)}$$

We don't often eat glucose in its simple form. More often, we eat complex carbohydrates that our bodies have to break down into individual glucose molecules before they can be used in cellular respiration. These complex carbohydrates are polymers, which form through condensation polymerisation reactions (figure 10.11). *Starch* and *cellulose* are two example of carbohydrates that are polymers composed of glucose monomers.

Figure 10.11: Two glucose monomers (a) undergo a condensation reaction to produce a section of a carbohydrate polymer (b). One molecule of water is produced for every two monomers that react.

- **Starch**

  Starch is used by plants to store excess glucose, and consists of long chains of glucose monomers. Potatoes are made up almost entirely of starch. This is why potatoes are such a good source of energy. Animals are also able to store glucose, but in this case it is stored as a compound called **glycogen**, rather than as starch.

- **Cellulose**

  Cellulose is also made up of chains of glucose molecules, but the bonding between the polymers is slightly different from that in starch. Cellulose is found in the cell walls of plants and is used by plants as a building material.

Interesting Fact

It is very difficult for animals to digest the cellulose in plants that they may have been feeding on. However, fungi and some protozoa are able to break down cellulose. Many animals, including termites and cows, use these organisms to break cellulose down into glucose, which they can then use more easily.

10.6.2 Proteins

Proteins are an incredibly important part of any cell, and they carry out a number of functions such as support, storage and transport within the body. The monomers of proteins are called **amino acids**. An amino acid is an organic molecule that contains a carboxyl and an amino group, as well as a carbon side chain. The carbon side chain varies from one amino acid to the next, and is sometimes simply represented by the letter 'R' in a molecule's structural formula. Figure 10.12 shows some examples of different amino acids.

Figure 10.12: Three amino acids: glycine, alanine and serine

Although each of these amino acids has the same basic structure, their side chains ('R' groups) are different. In the amino acid glycine, the side chain consists only of a hydrogen atom, while alanine has a *methyl* side chain. The 'R' group in serine is CH_2 - OH. Amongst other things, the side chains affect whether the amino acid is *hydrophilic* (attracted to water) or *hydrophobic* (repelled by water). If the side chain is *polar*, then the amino acid is hydrophilic, but if the side chain is *non-polar* then the amino acid is hydrophobic. Glycine and alanine both have non-polar side chains, while serine has a polar side chain.

Extension: Charged regions in an amino acid

In an amino acid, the amino group acts as a base because the nitrogen atom has a pair of unpaired electrons which it can use to bond to a hydrogen ion. The amino group therefore attracts the hydrogen ion from the carboxyl group, and ends up having a charge of +1. The carboxyl group from which the hydrogen ion has been taken then has a charge of -1. The amino acid glycine can therefore also be represented as shown in the figure below.

glycine

When two amino acid monomers are close together, they may be joined to each other by **peptide bonds** (figure 10.13) to form a **polypeptide** chain. . The reaction is a condensation reaction. Polypeptides can vary in length from a few amino acids to a thousand or more. The polpeptide chains are then joined to each other in different ways to form a **protein**. It is the sequence of the amino acids in the polymer that gives a protein its particular properties.

The sequence of the amino acids in the chain is known as the protein's **primary structure**. As the chain grows in size, it begins to twist, curl and fold upon itself. The different parts of the polypeptide are held together by hydrogen bonds, which form between hydrogen atoms in one part of the chain and oxygen or nitrogen atoms in another part of the chain. This is known as the **secondary structure** of the protein. Sometimes, in this coiled helical structure, bonds may form between the side chains (R groups) of the amino acids. This results in even more irregular contortions of the protein. This is called the **tertiary structure** of the protein.

Figure 10.13: Two amino acids (glycine and alanine) combine to form part of a polypeptide chain. The amino acids are joined by a peptide bond between a carbon atom of one amino acid and a nitrogen atom of the other amino acid.

Interesting Fact

There are twenty different amino acids that exist. All cells, both plant and animal, build their proteins from only twenty amino acids. At first, this seems like a very small number, especially considering the huge number of different

proteins that exist. However, if you consider that most proteins are made up of polypeptide chains that contain at least 100 amino acids, you will start to realise the endless possible combinations of amino acids that are available.

The functions of proteins

Proteins have a number of functions in living organisms.

- *Structural proteins* such as collagen in animal connective tissue and keratin in hair, horns and feather quills, all provide support.
- *Storage proteins* such as albumin in egg white provide a source of energy. Plants store proteins in their seeds to provide energy for the new growing plant.
- *Transport proteins* transport other substances in the body. Haemoglobin in the blood for example, is a protein that contains iron. Haemoglobin has an affinity (attraction) for oxygen and so this is how oxygen is transported around the body in the blood.
- *Hormonal proteins* coordinate the body's activities. Insulin for example, is a hormonal protein that controls the sugar levels in the blood.
- *Enzymes* are chemical catalysts and speed up chemical reactions. Digestive enzymes such as salivary amylase in your saliva, help to break down polymers in food. Enzymes play an important role in all cellular reactions such as respiration, photosynthesis and many others.

Activity :: Research Project : Macromolecules in our daily diet

1. In order to keep our bodies healthy, it is important that we eat a balanced diet with the right amounts of carbohydrates, proteins and fats. Fats are an important source of energy, they provide insulation for the body, and they also provide a protective layer around many vital organs. Our bodies also need certain essential vitamins and minerals. Most food packaging has a label that provides this information.
   Choose a number of different food items that you eat. Look at the food label for each, and then complete the following table:

Food	Carbohydrates (%)	Proteins (%)	Fats (%)

   (a) Which food type contains the largest proportion of protein?
   (b) Which food type contains the largest proportion of carbohydrates?
   (c) Which of the food types you have listed would you consider to be the 'healthiest'? Give a reason for your answer.

2. In an effort to lose weight, many people choose to *diet*. There are many diets on offer, each of which is based on particular theories about how to lose weight most effectively. Look at the list of diets below:
   - Vegetarian diet
   - Low fat diet
   - Atkin's diet

- Weight Watchers

For each of these diets, answer the following questions:

(a) What theory of weight loss does each type of diet propose?
(b) What are the *benefits* of the diet?
(c) What are the potential *problems* with the diet?

Exercise: Carbohydrates and proteins

1. Give the structural formula for each of the following:
   (a) A polymer chain, consisting of three glucose molecules.
   (b) A polypeptide chain, consisting of two molecules of alanine and one molecule of serine.
2. Write balanced equations to show the polymerisation reactions that produce the polymers described above.
3. The following polypeptide is the end product of a polymerisation reaction:

   (a) Give the structural formula of the monomers that make up the polypeptide.
   (b) On the structural formula of the first monomer, label the amino group and the carboxyl group.
   (c) What is the chemical formula for the carbon side chain in the second monomer?
   (d) Name the bond that forms between the monomers of the polypeptide.

10.6.3 Nucleic Acids

You will remember that we mentioned earlier that each protein is different because of its unique sequence of amino acids. But what controls how the amino acids arrange themselves to form the specific proteins that are needed by an organism? This task is for the **gene**. A gene contains DNA (deoxyribonucleic acid) which is a polymer that belongs to a class of compounds called the **nucleic acids**. DNA is the genetic material that organisms inherit from their parents. It is DNA that provides the genetic coding that is needed to form the specific proteins that an organism needs. Another nucleic acid is RNA (ribonucleic acid).

The DNA polymer is made up of monomers called **nucleotides**. Each nucleotide has three parts: a sugar, a phosphate and a nitrogenous base. The diagram in figure 10.14 may help you to understand this better.

Figure 10.14: Nucleotide monomers make up the DNA polymer

There are five different nitrogenous bases: adenine (A), guanine (G), cytosine (C), thymine (T) and uracil (U). It is the sequence of the nitrogenous bases in a DNA polymer that will determine the genetic code for that organism. Three consecutive nitrogenous bases provide the coding for one amino acid. So, for example, if the nitrogenous bases on three nucleotides are *uracil*, *cytosine* and *uracil* (in that order), one **serine** amino acid will become part of the polypeptide chain. The polypeptide chain is built up in this way until it is long enough (and with the right amino acid sequence) to be a protein. Since proteins control much of what happens in living organisms, it is easy to see how important nucleic acids are as the starting point of this process.

Interesting Fact

A single defect in even one nucleotide, can be devastating to an organism. One example of this is a disease called **sickle cell anaemia**. Because of one wrong nucletide in the genetic code, the body produces a protein called **sickle haemoglobin**. Haemoglobin is the protein in red blood cells that helps to transport oxygen around the body. When sickle haemoglobin is produced, the red blood cells change shape. This process damages the red blood cell membrane, and can cause the cells to become stuck in blood vessels. This then means that the red blood cells, whcih are carrying oxygen, can't get to the tissues where they are needed. This can cause serious organ damage. Individuals who have sickle cell anaemia generally have a lower life expectancy.

Table 10.2 shows some other examples of genetic coding for different amino acids.

Exercise: Nucleic acids

Table 10.2: Nitrogenouse base sequences and the corresponding amino acid

Nitrogenous base sequence	Amino acid
UUU	Phenylalanine
CUU	Leucine
UCU	Serine
UAU	Tyrosine
UGU	Cysteine
GUU	Valine
GCU	Alanine
GGU	Glycine

1. For each of the following, say whether the statement is **true** or **false**. If the statement is *false*, give a reason for your answer.
   (a) Deoxyribonucleic acid (DNA) is an example of a *polymer* and a nucleotide is an example of a *monomer*.
   (b) Thymine and uracil are examples of nucleotides.
   (c) A person's DNA will determine what proteins their body will produce, and therefore what characteristics they will have.
   (d) An amino acid is a protein monomer.
   (e) A polypeptide that consists of five amino acids, will also contain five nucleotides.
2. For each of the following sequences of nitrogenous bases, write down the amino acid/s that will be part of the polypeptide chain.
   (a) UUU
   (b) UCUUUU
   (c) GGUUAUGUUGGU
3. A polypeptide chain consists of three amino acids. The sequence of nitrogenous bases in the nucleotides of the DNA is GCUGGUGCU. Give the structural formula of the polypeptide.

10.7 Summary

- A **polymer** is a macromolecule that is made up of many repeating structural units called **monomers** which are joined by covalent bonds.
- Polymers that contain carbon atoms in the main chain are called **organic polymers**.
- Organic polymers can be divided into **natural organic polymers** (e.g. natural rubber) or **synthetic organic polymers** (e.g. polystyrene).
- The polymer **polyethene** for example, is made up of many ethene monomers that have been joined into a polymer chain.
- Polymers form through a process called **polymerisation**.
- Two examples of polymerisation reactions are **addition** and **condensation** reactions.
- An **addition reaction** occurs when unsaturated monomers (e.g. alkenes) are added to each other one by one. The breaking of a double bond between carbon atoms in the monomer, means that a bond can form with the next monomer. The polymer **polyethene** is formed through an addition reaction.

- In a **condensation reaction**, a molecule of water is released as a product of the reaction. The water molecule is made up of atoms that have been lost from each of the monomers. Polyesters and nylon are polymers that are produced through a condensation reaction.
- The **chemical properties** of polymers (e.g. tensile strength and melting point) are determined by the types of atoms in the polymer, and by the strength of the bonds between adjacent polymer chains. The stronger the bonds, the greater the strength of the polymer, and the higher its melting point.
- One group of synthetic organic polymers, are the **plastics**.
- **Polystyrene** is a plastic that is made up of styrene monomers. Polystyrene is used a lot in packaging.
- **Polyvinyl chloride** (PVC) consists of vinyl chloride monomers. PVC is used to make pipes and flooring.
- **Polyethene**, or **polyethylene**, is made from ethene monomers. Polyethene is used to make film wrapping, plastic bags, electrical insulation and bottles.
- **Polytetrafluoroethylene** is used in non-stick frying pans and electrical insulation.
- A **thermoplastic** can be heated and melted to a liquid. It freezes to a brittle, glassy state when cooled very quickly. Examples of thermoplastics are polyethene and PVC.
- A **thermoset** plastic cannot be melted or re-shaped once formed. Examples of thermoset plastics are vulcanised rubber and melanine.
- It is not easy to **recycle** all plastics, and so they create environmental problems.
- Some of these **environmental problems** include issues of waste disposal, air pollution and recycling.
- A **biological macromolecule** is a polymer that occurs naturally in living organisms.
- Examples of biological macromolecules include **carbohydrates** and **proteins**, both of which are essential for life to survive.
- Carbohydrates include the **sugars** and their polymers, and are an important source of **energy** in living organisms.
- **Glucose** is a carbohydrate monomer. Glucose is the molecule that is needed for **photosynthesis** in plants.
- The glucose monomer is also a building block for carbohydrate polymers such as **starch**, **glycogen** and **cellulose**.
- **Proteins** have a number of important functions. These include their roles in structures, transport, storage, hormonal proteins and enzymes.
- A protein consists of monomers called **amino acids**, which are joined by **peptide bonds**.
- A protein has a **primary**, **secondary** and **tertiary** structure.
- An amino acid is an organic molecule, made up of a **carboxyl** and an **amino** group, as well as a carbon **side chain** of varying lengths.
- It is the **sequence** of amino acids that determines the nature of the protein.
- It is the **DNA** of an organism that determines the order in which amino acids combine to make a protein.
- DNA is a **nucleic acid**. DNA is a polymer, and is made up of monomers called **nucleotides**.
- Each nucleotide consists of a **sugar**, a **phosphate** and a **nitrogenous base**. It is the sequence of the nitrogenous bases that provides the 'code' for the arrangement of the amino acids in a protein.

Exercise: Summary exercise

1. Give one word for each of the following descriptions:
   (a) A chain of monomers joined by covalent bonds.
   (b) A polymerisation reaction that produces a molecule of water for every two monomers that bond.
   (c) The bond that forms between an alcohol and a carboxylic acid monomer during a polymerisation reaction.
   (d) The name given to a protein monomer.
   (e) A six-carbon sugar monomer.
   (f) The monomer of DNA, which determines the sequence of amino acids that will make up a protein.
2. For each of the following questions, choose the one correct answer from the list provided.
   (a) A polymer is made up of monomers, each of which has the formula $CH_2{=}CHCN$. The formula of the polymer is:
      i. -$(CH_2{=}CHCN)_n$-
      ii. -$(CH_2\text{-}CHCN)_n$-
      iii. -$(CH\text{-}CHCN)_n$-
      iv. -$(CH_3\text{-}CHCN)_n$-
   (b) A polymer has the formula -$[CO(CH_2)_4CO\text{-}NH(CH_2)6NH]_n$-. Which of the following statements is **true**?
      i. The polymer is the product of an addition reaction.
      ii. The polymer is a polyester.
      iii. The polymer contains an amide linkage.
      iv. The polymer contains an ester linkage.
   (c) Glucose...
      i. is a monomer that is produced during cellular respiration
      ii. is a sugar polymer
      iii. is the monomer of starch
      iv. is a polymer produced during photosynthesis
3. The following monomers are involved in a polymerisation reaction:

   (a) Give the structural formula of the polymer that is produced.
   (b) Is the reaction an addition or condensation reaction?
   (c) To what group of organic compounds do the two monomers belong?
   (d) What is the name of the monomers?
   (e) What type of bond forms between the monomers in the final polymer?
4. The table below shows the melting point for three plastics. Suggest a reason why the melting point of PVC is *higher* than the melting point for polyethene, but *lower* than that for polyester.

Plastic	**Melting point** (^{0}C)
Polyethene	105 - 115
PVC	212
Polyester	260

5. An amino acid has the formula $H_2NCH(CH_2CH_2SCH_3)COOH$.
   (a) Give the structural formula of this amino acid.
   (b) What is the chemical formula of the carbon side chain in this molecule?
   (c) Are there any peptide bonds in this molecule? Give a reason for your answer.

Part III

Chemical Change

Chapter 11

Physical and Chemical Change - Grade 10

Matter is all around us. The desks we sit at, the air we breathe and the water we drink, are all examples of matter. But matter doesn't always stay the same. It can change in many different ways. In this chapter, we are going to take a closer look at **physical** and **chemical** changes that occur in matter.

11.1 Physical changes in matter

A **physical change** is one where the particles of the substances that are involved in the change are not broken up in any way. When water is heated for example, the temperature and energy of the water molecules increases and the liquid water evaporates to form water vapour. When this happens, some kind of change has taken place, but the molecular structure of the water has not changed. This is an example of a *physical change*.

$$H_2O(l) \rightarrow H_2O(g)$$

Conduction (the transfer of energy through a material) is another example of a physical change. As energy is transferred from one material to another, the *energy* of each material is changed, but not its chemical makeup. Dissolving one substance in another is also a physical change.

Definition: Physical change
A change that can be seen or felt, but that doesn't involve the break up of the particles in the reaction. During a physical change, the *form* of matter may change, but not its *identity*. A change in temperature is an example of a physical change.

There are some important things to remember about physical changes in matter:

- *Arrangement of particles*

  When a physical change occurs, the particles (e.g. atoms, molecules) may re-arrange themselves without actually breaking up in any way. In the example of evaporation that we used earlier, the water molecules move further apart as their temperature (and therefore energy) increases. The same would be true if ice were to melt. In the solid phase, water molecules are packed close together in a very ordered way, but when the ice is heated, the molecules overcome the forces holding them together and they move apart. Once again, the particles have re-arranged themselves, but have not broken up.

$$H_2O(s) \rightarrow H_2O(l)$$

Figure 11.1: The arrangement of water molecules in the three phases of matter

Figure 11.1 shows this more clearly. In each phase of water, the water molecule itself stays the same, but the way the molecules are arranged has changed.

In a physical change, the total mass, the number of atoms and the number of molecules will always stay the same.

- *Energy changes*

  Energy changes may take place when there is a physical change in matter, but these energy changes are normally smaller than the energy changes that take place during a chemical change.

- *Reversibility*

  Physical changes in matter are usually easier to reverse than chemical changes. Water vapour for example, can be changed back to liquid water if the temperature is lowered. Liquid water can be changed into ice by simply increasing the temperature, and so on.

11.2 Chemical Changes in Matter

When a **chemical change** takes place, new substances are formed in a chemical reaction. These new products may have very different properties from the substances that were there at the start of the reaction.

The breakdown of copper(II) chloride to form copper and chlorine is an example of chemical change. A simplified diagram of this reaction is shown in figure 11.2. In this reaction, the initial substance is copper(II) chloride but, once the reaction is complete, the products are copper and chlorine.

$$CuCl_2 \rightarrow Cu + Cl_2$$

Figure 11.2: The decomposition of copper(II) chloride to form copper and chlorine

Definition: Chemical change
The formation of new substances in a chemical reaction. One type of matter is changed into something different.

There are some important things to remember about chemical changes:

- *Arrangement of particles*

During a chemical change, the particles themselves are changed in some way. In the example of copper (II) chloride that was used earlier, the $CuCl_2$ molecules were split up into their component atoms. The number of particles will change because each one $CuCl_2$ molecule breaks down into one copper atom (Cu) and one chlorine molecule (Cl_2). However, what you should have noticed, is that the number of atoms of each element stays the same, as does the total mass of the atoms. This will be discussed in more detail in a later section.

- *Energy changes*

  The energy changes that take place during a chemical reaction are much greater than those that take place during a physical change in matter. During a chemical reaction, energy is used up in order to break bonds, and then energy is released when the new product is formed. This will be discussed in more detail in section **??**.

- *Reversibility*

  Chemical changes are far more difficult to reverse than physical changes.

Two types of chemical reactions are **decomposition reactions** and **synthesis reactions**.

11.2.1 Decomposition reactions

A **decomposition reaction** occurs when a chemical compound is broken down into elements or smaller compounds. The generalised equation for a decomposition reaction is:

$$AB \rightarrow A + B$$

One example of such a reaction is the decomposition of hydrogen peroxide (figure 11.3) to form hydrogen and oxygen according to the following equation:

$$2H_2O_2 \rightarrow 2H_2O + O_2$$

Figure 11.3: The decomposition of H_2O_2 to form H_2O and O_2

The decomposition of mercury (II) oxide is another example.

Activity :: Experiment : The decomposition of mercury (II) oxide
Aim:

To observe the decomposition of mercury (II) oxide when it is heated.
Note: Because this experiment involves mercury, which is a poisonous substance, it should be done in a fume cupboard, and all the products of the reaction must be very carefully disposed of.
Apparatus:

Mercury (II) oxide (an orange-red product); two test tubes; a large beaker; stopper and delivery tube; Bunsen burner; wooden splinter.

Method:

1. Put a small amount of mercury (II) oxide in a test tube and heat it gently over a Bunsen burner. Then allow it to cool. What do you notice about the colour of the mercury (II) oxide?
2. Heat the test tube again, and note what happens. Do you notice anything on the walls of the test tube? Record these observations.
3. Test for the presence of oxygen using a glowing splinter.

Results:

- During the first heating of mercury (II) oxide, the only change that took place was a change in colour from orange-red to black and then back to its original colour.
- When the test tube was heated again, deposits of mercury formed on the inner surface of the test tube. What colour is this mercury?
- The glowing splinter burst into flame when it was placed in the test tube, meaning that oxygen is present.

Conclusions:

When mercury is heated, it decomposes to form mercury and oxygen. The chemical decomposition reaction that takes place can be written as follows:

$$2HgO \rightarrow 2Hg + O_2$$

11.2.2 Synthesis reactions

During a **synthesis reaction**, a new product is formed from smaller elements or compounds. The generalised equation for a synthesis reaction is as follows:

$$A + B \rightarrow AB$$

One example of a synthesis reaction is the burning of magnesium in oxygen to form magnesium oxide. The equation for the reaction is:

$$2Mg + O_2 \rightarrow 2MgO$$

Figure 11.4 shows the chemical changes that take place at a microscopic level during this chemical reaction.

Figure 11.4: The synthesis of magnesium oxide (MgO) from magnesium and oxygen

---

Activity :: Experiment : Chemical reactions involving iron and sulfur

Aim:

To demonstrate the synthesis of iron sulfide from iron and sulfur.

Apparatus:

5.6 g iron filings and 3.2 g powdered sulfur; porcelain dish; test tube; bunsen burner

Method:

1. Before you carry out the experiment, write a balanced equation for the reaction you expect will take place.
2. Measure the quantity of iron and sulfur that you need and mix them in a porcelain dish.
3. Take some of this mixture and place it in the test tube. The test tube should be about 1/3 full.
4. This reaction should ideally take place in a fume cupboard. Heat the test tube containing the mixture over the Bunsen burner. Increase the heat if no reaction takes place. Once the reaction begins, you will need to remove the test tube from the flame. Record your observations.

5. Wait for the product to cool before breaking the test tube with a hammer. Make sure that the test tube is rolled in paper before you do this, otherwise the glass will shatter everywhere and you may be hurt.
6. What does the product look like? Does it look anything like the original reactants? Does it have any of the properties of the reactants (e.g. the magnetism of iron)?

Results:

- After you removed the test tube from the flame, the mixture glowed a bright red colour. The reaction is exothermic and *produces energy*.
- The product, iron sulfide, is a dark colour and does not share any of the properties of the original reactants. It is an entirely new product.

Conclusions:

A synthesis reaction has taken place. The equation for the reaction is:

$$Fe + S \rightarrow FeS$$

Activity :: Investigation : Physical or chemical change?

Apparatus:

Bunsen burner, 4 test tubes, a test tube rack and a test tube holder, small spatula, pipette, magnet, a birthday candle, NaCl (table salt), 0.1M $AgNO_3$, 6M HCl, magnesium ribbon, iron filings, sulfur.

Method:

1. Place a small amount of wax from a birthday candle into a test tube and heat it over the bunsen burner until it melts. Leave it to cool.
2. Add a small spatula of NaCl to 5 ml water in a test tube and shake. Then use the pipette to add 10 drops of $AgNO_3$ to the sodium chloride solution.
3. Take a 5 cm piece of magnesium ribbon and tear it into 1 cm pieces. Place two of these pieces into a test tube and add a few drops of 6M HCl. NOTE: Be very careful when you handle this acid because it can cause major burns.
4. Take about 0.5 g iron filings and 0.5 g sulfur. Test each substance with a magnet. Mix the two samples in a test tube, and run a magnet alongside the outside of the test tube.
5. Now heat the test tube that contains the iron and sulfur. What changes do you see? What happens now, if you run a magnet along the outside of the test tube?
6. In each of the above cases, record your observations.

Questions:

Decide whether each of the following changes are physical or chemical and give a reason for your answer in each case. Record your answers in the table below:

Description	**Physical or chemical change**	**Reason**
melting candle wax		
dissolving NaCl		
mixing NaCl with $AgNO_3$		
tearing magnesium ribbon		
adding HCl to magnesium ribbon		
mixing iron and sulfur		
heating iron and sulfur		

11.3 Energy changes in chemical reactions

All reactions involve some change in energy. During a *physical* change in matter, such as the evaporation of liquid water to water vapour, the energy of the water molecules increases. However, the change in energy is much smaller than in chemical reactions.

When a chemical reaction occurs, some bonds will *break*, while new bonds may *form*. Energy changes in chemical reactions result from the breaking and forming of bonds. For bonds to *break*, energy must be *absorbed*. When new bonds *form*, energy will be *released* because the new product has a lower energy than the 'inbetween' stage of the reaction when the bonds in the reactants have just been broken.

In some reactions, the energy that must be *absorbed* to break the bonds in the reactants, is less than the total energy that is *released* when new bonds are formed. This means that in the overall reaction, energy is *released*. This type of reaction is known as an **exothermic** reaction. In other reactions, the energy that must be *absorbed* to break the bonds in the reactants, is more than the total energy that is *released* when new bonds are formed. This means that in the overall reaction, energy must be *absorbed* from the surroundings. This type of reaction is known as an **endothermic** reaction. In the earlier part of this chapter, most decomposition reactions were endothermic, and heating was needed for the reaction to occur. Most of the synthesis reactions were exothermic, meaning that energy was given off in the form of heat or light.

More simply, we can describe the energy changes that take place during a chemical reaction as:

Total energy absorbed to break bonds - Total energy released when new bonds form

So, for example, in the reaction...

$$2Mg + O_2 \rightarrow 2MgO$$

Energy is needed to break the O-O bonds in the oxygen molecule so that new Mg-O bonds can be formed, and energy is released when the product (MgO) forms.

Despite all the energy changes that seem to take place during reactions, it is important to remember that energy cannot be created or destroyed. Energy that enters a system will have come from the surrounding environment, and energy that leaves a system will again become part of that environment. This principle is known as the principle of **conservation of energy.**

Definition: Conservation of energy principle
Energy cannot be created or destroyed. It can only be changed from one form to another.

Chemical reactions may produce some very visible, and often violent, changes. An explosion, for example, is a sudden increase in volume and release of energy when high temperatures are generated and gases are released. For example, NH_4NO_3 can be heated to generate nitrous oxide. Under these conditions, it is highly sensitive and can detonate easily in an explosive exothermic reaction.

11.4 Conservation of atoms and mass in reactions

The total mass of all the substances taking part in a chemical reaction is conserved during a chemical reaction. This is known as the **law of conservation of mass**. The total number of **atoms** of each element also remains the same during a reaction, although these may be arranged differently in the products.

We will use two of our earlier examples of chemical reactions to demonstrate this:

- The decomposition of hydrogen peroxide into water and oxygen

$$2H_2O_2 \rightarrow 2H_2O + O_2$$

Left hand side of the equation

Total atomic mass = (4 × 1) + (4 × 16) = 68 u

Number of atoms of each element = (4 × H) + (4 × O)

Right hand side of the equation

Total atomic mass = (4 × 1) + (2 × 16) + (2 × 16) = 68 u

Number of atoms of each element = (4 × H) + (4 × O)

Both the atomic mass and the number of atoms of each element are conserved in the reaction.

- The synthesis of magnesium and oxygen to form magnesium oxide

$$2Mg + O_2 \rightarrow 2MgO$$

Left hand side of the equation

Total atomic mass = (2 × 24.3) + (2 × 16) = 80.6 u

Number of atoms of each element = (2 × Mg) + (2 × O)

Right hand side of the equation

Total atomic mass = (2 × 24.3) + (2 × 16) = 80.6 u

Number of atoms of each element = (2 × Mg) + (2 × O)

Both the atomic mass and the number of atoms of each element are conserved in the reaction.

Activity :: Demonstration : The conservation of atoms in chemical reactions

Materials:

- Coloured marbles or small balls to represent atoms. Each colour will represent a different element.

- Prestik

Method:

1. Choose a reaction from any that have been used in this chapter or any other *balanced* chemical reaction that you can think of. To help to explain this activity, we will use the decomposition reaction of calcium carbonate to produce carbon dioxide and calcium oxide.

$$CaCO_3 \rightarrow CO_2 + CaO$$

2. Stick marbles together to represent the reactants and put these on one side of your table. In this example you may for example join one red marble (calcium), one green marble (carbon) and three yellow marbles (oxygen) together to form the molecule calcium carbonate ($CaCO_3$).
3. Leaving your reactants on the table, use marbles to make the product molecules and place these on the other side of the table.
4. Now count the number of atoms on each side of the table. What do you notice?
5. Observe whether there is any difference between the molecules in the reactants and the molecules in the products.

Discussion

You should have noticed that the number of atoms in the reactants is the same as the number of atoms in the product. The number of atoms is conserved during the reaction. However, you will also see that the molecules in the reactants and products is not the same. The *arrangement of atoms* is not conserved during the reaction.

11.5 Law of constant composition

In any given chemical compound, the elements always combine in the same proportion with each other. This is the **law of constant proportions**.

The **law of constant composition** says that, in any particular chemical compound, all samples of that compound will be made up of the same elements in the same proportion or ratio. For example, any water molecule is always made up of two hydrogen atoms and one oxygen atom in a 2:1 ratio. If we look at the relative masses of oxygen and hydrogen in a water molecule, we see that 94% of the mass of a water molecule is accounted for by oxygen, and the remaining 6% is the mass of hydrogen. This mass proportion will be the same for any water molecule.

This does not mean that hydrogen and oxygen always combine in a 2:1 ratio to form H_2O. Multiple proportions are possible. For example, hydrogen and oxygen may combine in different proportions to form H_2O_2 rather than H_2O. In H_2O_2, the H:O ratio is 1:1 and the mass ratio of hydrogen to oxygen is 1:16. This will be the same for any molecule of hydrogen peroxide.

11.6 Volume relationships in gases

In a chemical reaction between gases, the relative volumes of the gases in the reaction are present in a ratio of small whole numbers if all the gases are at the same temperature and pressure. This relationship is also known as **Gay-Lussac's Law**.

For example, in the reaction between hydrogen and oxygen to produce water, two volumes of H_2 react with 1 volume of O_2 to produce 2 volumes of H_2O.

$$2H_2 + O_2 \rightarrow 2H_2O$$

In the reaction to produce ammonia, one volume of nitrogen gas reacts with three volumes of hydrogen gas to produce two volumes of ammonia gas.

$$N_2 + 3H_2 \rightarrow 2NH_3$$

This relationship will also be true for all other chemical reactions.

11.7 Summary

- Matter does not stay the same. It may undergo physical or chemical changes
- A **physical change** means that the form of matter may change, but not its identity. For example, when water evaporates, the energy and the arrangement of water molecules will change, but not the structure of the water molecule itself.
- During a physical change, the **arrangement of particles** may change but the mass, number of atoms and number of molecules will stay the same.
- Physical changes involve small changes in **energy**, and are easily reversible.
- A chemical change occurs when one form of matter changes into something else. A chemical reaction involves the formation of new substances with **different properties**. For example, carbon dioxide reacts with water to form carbonic acid.

$$CO_2 + H_2O \rightarrow H_2CO_3$$

- A chemical change may involve a **decomposition** or **synthesis** reaction. During chemical change, the mass and number of atoms is conserved, but the number of molecules is not always the same.
- Chemical reactions involve larger changes in energy. During a reaction, energy is needed to break bonds in the reactants, and energy is released when new products form. If the energy released is greater than the energy absorbed, then the reaction is exothermic. If the energy released is less than the energy absorbed, then the reaction is endothermic. These chemical reactions are not easily reversible.
- Decomposition reactions are usually **endothermic** and synthesis reactions are usually **exothermic**.
- The **law of conservation of mass** states that the total mass of all the substances taking part in a chemical reaction is conserved and the number of atoms of each element in the reaction does not change when a new product is formed.
- The **conservation of energy principle** states that energy cannot be created or destroyed, it can only change from one form to another.
- The **law of constant composition** states that in any particular compound, all samples of that compound will be made up of the same elements in the same proportion or ratio.
- **Gay-Lussac's Law** states that in a chemical reaction between gases, the relative volumes of the gases in the reaction are present in a ratio of small whole numbers if all the gases are at the same temperature and pressure.

Exercise: Summary exercise

1. Complete the following table by saying whether each of the descriptions is an example of a physical or chemical change:

Description	**Physical or chemical**
hot and cold water mix together	
milk turns sour	
a car starts to rust	
food digests in the stomach	
alcohol disappears when it is placed on your skin	
warming food in a microwave	
separating sand and gravel	
fireworks exploding	

2. For each of the following reactions, say whether it is an example of a synthesis or decomposition reaction:
   (a) $(NH_4)_2CO_3 \rightarrow 2NH_3 + CO_2 + H_2O$
   (b) $4Fe + 3O_2 \rightarrow 2Fe_2O_3$
   (c) $N_2(g) + 3H_2(g) \rightarrow 2NH_3$
   (d) $CaCO_3(s) \rightarrow CaO + CO_2$
3. For the following equation:

$$CaCO_3 \rightarrow CO_2 + CaO$$

   Show that the 'law of conservation of mass' applies.

Chapter 12

Representing Chemical Change - Grade 10

As we have already mentioned, a number of changes can occur when elements react with one another. These changes may either be *physical* or *chemical*. One way of representing these changes is through **balanced chemical equations**. A chemical equation describes a chemical reaction by using symbols for the elements involved. For example, if we look at the reaction between iron (Fe) and sulfur (S) to form iron sulfide (FeS), we could represent these changes either in words or using chemical symbols:

$$\text{iron} + \text{sulfur} \rightarrow \text{iron sulfide}$$

$$\text{or}$$

$$Fe + S \rightarrow FeS$$

Another example would be:

$$\text{ammonia} + \text{oxygen} \rightarrow \text{nitric oxide} + \text{water}$$

$$\text{or}$$

$$4NH_3 + 5O_2 \rightarrow 4NO + 6H_2O$$

Compounds on the left of the arrow are called the **reactants** and these are needed for the reaction to take place. In this equation, the reactants are ammonia and oxygen. The compounds on the right are called the **products** and these are what is formed from the reaction.

In order to be able to write a balanced chemical equation, there are a number of important things that need to be done:

1. Know the chemical symbols for the elements involved in the reaction
2. Be able to write the chemical formulae for different reactants and products
3. Balance chemical equations by understanding the laws that govern chemical change
4. Know the state symbols for the equation

We will look at each of these steps separately in the next sections.

12.1 Chemical symbols

It is very important to know the chemical symbols for common elements in the Periodic Table so that you are able to write chemical equations and to recognise different compounds.

?

Exercise: Revising common chemical symbols

- Write down the chemical symbols and names of all the elements that you know.
- Compare your list with another learner and add any symbols and names that you don't have.
- Spend some time, either in class or at home, learning the symbols for at least the first twenty elements in the periodic table. You should also learn the symbols for other common elements that are not in the first twenty.
- Write a short test for someone else in the class and then exchange tests with them so that you each have the chance to answer one.

12.2 Writing chemical formulae

A **chemical formula** is a concise way of giving information about the atoms that make up a particular chemical compound. A chemical formula shows each element by its symbol, and also shows how many atoms of each element are found in that compound. The number of atoms (if greater than one) is shown as a subscript.

Examples:

CH_4 (methane)

Number of atoms: (1 x carbon) + (4 x hydrogen) = 5 atoms in one methane molecule

H_2SO_4 (sulfuric acid)

Number of atoms: (2 x hydrogen) + (1 x sulfur) + (4 x oxygen) = 7 atoms in one molecule of sulfuric acid

A chemical formula may also give information about how the atoms are arranged in a molecule if it is written in a particular way. A molecule of ethane, for example, has the chemical formula C_2H_6. This formula tells us how many atoms of each element are in the molecule, but doesn't tell us anything about how these atoms are arranged. In fact, each carbon atom in the ethane molecule is bonded to three hydrogen atoms. Another way of writing the formula for ethane is CH_3CH_3. The number of atoms of each element has not changed, but this formula gives us more information about how the atoms are arranged in relation to each other.

The slightly tricky part of writing chemical formulae comes when you have to work out the ratio in which the elements combine. For example, you may know that sodium (Na) and chlorine (Cl) react to form sodium chloride, but how do you know that in each molecule of sodium chloride there is only *one* atom of sodium for every *one* atom of chlorine? It all comes down to the **valency** of an atom or group of atoms. Valency is the number of bonds that an element can form with another element. Working out the chemical formulae of chemical compounds using their valency, will be covered in chapter 4. For now, we will use formulae that you already know.

12.3 Balancing chemical equations

12.3.1 The law of conservation of mass

In order to balance a chemical equation, it is important to understand the law of conservation of mass.

Definition: The law of conservation of mass
The mass of a closed system of substances will remain constant, regardless of the processes acting inside the system. Matter can change form, but cannot be created or destroyed. For any chemical process in a closed system, the mass of the reactants must equal the mass of the products.

In a chemical equation then, the **mass** of the reactants must be equal to the mass of the products. In order to make sure that this is the case, the number of **atoms** of each element in the reactants must be equal to the number of atoms of those same elements in the products. Some examples are shown below:

Example 1:

$$Fe + S \rightarrow FeS$$

Reactants

Atomic mass of reactants = 55.8 u + 32.1 u = 87.9 u

Number of atoms of each element in the reactants: (1 × Fe) and (1 × S)

Products

Atomic mass of product = 55.8 u + 32.1 u = 87.9 u

Number of atoms of each element in the products: (1 × Fe) and (1 × S)

Since the number of atoms of each element is the same in the reactants and in the products, we say that the equation is **balanced**.

Example 2:

$$H_2 + O_2 \rightarrow H_2O$$

Reactants

Atomic mass of reactants = (1 + 1) + (16 + 16) = 34 u

Number of atoms of each element in the reactants: (2 × H) and (2 × O)

Product

Atomic mass of product = (1 + 1 + 16) = 18 u

Number of atoms of each element in the products: (2 × H) and (1 × O)

Since the total atomic mass of the reactants and the products is not the same, and since there are more oxygen atoms in the reactants than there are in the product, the equation is **not balanced**.

Example 3:

$$NaOH + HCl \rightarrow NaCl + H_2O$$

Reactants

Atomic mass of reactants = (23 + 16 + 1) + (1 + 35.4) = 76.4 u

Number of atoms of each element in the reactants: (1 × Na) + (1 × O) + (2 × H) + (1 × Cl)

Products

Atomic mass of products = (23 + 35.4) + (1 + 1 + 16) = 76.4 u

Number of atoms of each element in the products: (1 × Na) + (1 × O) + (2 × H) + (1 × Cl)

Since the number of atoms of each element is the same in the reactants and in the products, we say that the equation is **balanced**.

We now need to find a way to balance those equations that are not balanced so that the number of atoms of each element in the reactants is the same as that for the products. This can be done by changing the **coefficients** of the molecules until the atoms on each side of the arrow are balanced. You will see later in chapter 13 that these coefficients tell us something about the **mole ratio** in which substances react. They also tell us about the volume relationship between gases in the reactants and products.

Important: Coefficients

Remember that if you put a number in front of a molecule, that number applies to the *whole* molecule. For example, if you write $2H_2O$, this means that there are 2 molecules of water. In other words, there are 4 hydrogen atoms and 2 oxygen atoms. If we write 3HCl, this means that there are 3 molecules of HCl. In other words there are 3 hydrogen atoms and 3 chlorine atoms in total. In the first example, 2 is the coefficient and in the second example, 3 is the coefficient.

12.3.2 Steps to balance a chemical equation

When balancing a chemical equation, there are a number of steps that need to be followed.

- STEP 1: Identify the reactants and the products in the reaction, and write their chemical formulae.
- STEP 2: Write the equation by putting the reactants on the left of the arrow, and the products on the right.
- STEP 3: Count the number of atoms of each element in the reactants and the number of atoms of each element in the products.
- STEP 4: If the equation is not balanced, change the coefficients of the molecules until the number of atoms of each element on either side of the equation balance.
- STEP 5: Check that the atoms are in fact balanced.
- STEP 6 (we will look at this a little later): Add any extra details to the equation e.g. phase.

Worked Example 49: Balancing chemical equations 1

Question: Balance the following equation:

$$Mg + HCl \rightarrow MgCl_2 + H_2$$

Answer

Step 1 : Because the equation has been written for you, you can move straight on to counting the number of atoms of each element in the reactants and products

Reactants: Mg = 1 atom; H = 1 atom and Cl = 1 atom
Products: Mg = 1 atom; H = 2 atoms and Cl = 2 atoms

Step 2 : Balance the equation

The equation is not balanced since there are 2 chlorine atoms in the product and only 1 in the reactants. If we add a coefficient of 2 to the HCl to increase the number of H and Cl atoms in the reactants, the equation will look like this:

$$Mg + 2HCl \rightarrow MgCl_2 + H_2$$

Step 3 : Check that the atoms are balanced

If we count the atoms on each side of the equation, we find the following:
Reactants: Mg = 1; H = 2; Cl = 2
Products: Mg = 1; H = 2; Cl = 2
The equation is balanced. The final equation is:

$$Mg + 2HCl \rightarrow MgCl_2 + H_2$$

Worked Example 50: Balancing chemical equations 2

Question: Balance the following equation:

$$CH_4 + O_2 \rightarrow CO_2 + H_2O$$

Answer

Step 1 : Count the number of atoms of each element in the reactants and products

Reactants: C = 1; H = 4; O = 2
Products: C = 1; H = 2; O = 3

Step 2 : Balance the equation

If we add a coefficient of 2 to H_2O, then the number of hydrogen atoms in the reactants will be 4, which is the same as for the reactants. The equation will be:

$$CH_4 + O_2 \rightarrow CO_2 + 2H_2O$$

Step 3 : Check that the atoms balance

Reactants: C = 1; H = 4; O = 2
Products: C = 1; H = 4; O = 4
You will see that, although the number of *hydrogen* atoms now balances, there are more oxygen atoms in the products. You now need to repeat the previous step. If we put a coefficient of 2 in front of O_2, then we will increase the number of oxygen atoms in the reactants by 2. The new equation is:

$$CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O$$

When we check the number of atoms again, we find that the number of atoms of each element in the reactants is the same as the number in the products. The equation is now balanced.

Worked Example 51: Balancing chemical equations 3

Question: Nitrogen gas reacts with hydrogen gas to form ammonia. Write a balanced chemical equation for this reaction.

Answer
Step 1 : Identify the reactants and the products, and write their chemical formulae
The reactants are nitrogen (N_2) and hydrogen (H_2), and the product is ammonia (NH_3).

Step 2 : Write the equation so that the reactants are on the left and products on the right of the arrow
The equation is as follows:

$$N_2 + H_2 \rightarrow NH_3$$

Step 3 : Count the atoms of each element in the reactants and products
Reactants: N = 2; H = 2
Products: N = 1; H = 3

Step 4 : Balance the equation
In order to balance the number of nitrogen atoms, we could rewrite the equation as:

$$N_2 + H_2 \rightarrow 2NH_3$$

Step 5 : Check that the atoms are balanced
In the above equation, the nitrogen atoms now balance, but the hydrogen atoms don't (there are 2 hydrogen atoms in the reactants and 6 in the product). If we put a coefficient of 3 in front of the hydrogen (H_2), then the hydrogen atoms and the nitrogen atoms balance. The final equation is:

$$N_2 + 3H_2 \rightarrow 2NH_3$$

Worked Example 52: Balancing chemical equations 4

Question: In our bodies, sugar ($C_6H_{12}O_6$) reacts with the oxygen we breathe in to produce carbon dioxide, water and energy. Write the balanced equation for this reaction.

Answer
Step 1 : Identify the reactants and products in the reaction, and write their chemical formulae.
Reactants: sugar ($C_6H_{12}O_6$) and oxygen (O_2)
Products: carbon dioxide (CO_2) and water (H_2O)

Step 2 : Write the equation by putting the reactants on the left of the arrow, and the products on the right

$$C_6H_{12}O_6 + O_2 \rightarrow CO_2 + H_2O$$

Step 3 : Count the number of atoms of each element in the reactants and the number of atoms of each element in the products

Reactants: C=6; H=12; O=8;
Products: C=1; H=2; O=3;

Step 4 : Change the coefficents of the molecules until the number of atoms of each element on either side of the equation balance.
It is easier to start with carbon as it only appears once on each side. If we add a 6 in front of CO_2, the equation looks like this:

$$C_6H_{12}O_6 + O_2 \rightarrow 6CO_2 + H_2O$$

Reactants: C=6; H=12; O=8;
Products: C=6; H=2; O=13;

Step 5 : Change the coefficients again to try to balance the equation.
Let's try to get the number of hydrogens the same this time.

$$C_6H_{12}O_6 + O_2 \rightarrow 6CO_2 + 6H_2O$$

Reactants: C=6; H=12; O=8;
Products: C=6; H=12; O=18;

Step 6 : Now we just need to balance the oxygen atoms.

$$C_6H_{12}O_6 + 6O_2 \rightarrow 6CO_2 + 6H_2O$$

Reactants: C=6; H=12; O=18;
Products: C=6; H=12; O=18;

?

Exercise: Balancing simple chemical equations
Balance the following equations:

1. Hydrogen fuel cells are extremely important in the development of alternative energy sources. Many of these cells work by reacting hydrogen and oxygen gases together to form water, a reaction which also produces electricity. Balance the following equation:
$$H_2(g) + O_2(g) \rightarrow H_2O(l)$$
2. The synthesis of ammonia (NH_3), made famous by the German chemist Fritz Haber in the early 20th century, is one of the most important reactions in the chemical industry. Balance the following equation used to produce ammonia:
$$N_2(g) + H_2(g) \rightarrow NH_3(g)$$
3. $Mg + P_4 \rightarrow Mg_3P_2$
4. $Ca + H_2O \rightarrow Ca(OH)_2 + H_2$
5. $CuCO_3 + H_2SO_4 \rightarrow CuSO_4 + H_2O + CO_2$
6. $CaCl_2 + Na_2CO_3 \rightarrow CaCO_3 + NaCl$
7. $C_{12}H_{22}O_{11} + O_2 \rightarrow CO_2 + H_2O$
8. Barium chloride reacts with sulphuric acid to produce barium sulphate and hydrochloric acid.
9. Ethane (C_2H_6) reacts with oxygen to form carbon dioxide and steam.
10. Ammonium carbonate is often used as a smelling salt. Balance the following reaction for the decomposition of ammonium carbonate:
$$(NH_4)_2CO_3(s) \rightarrow NH_3(aq) + CO_2(g) + H_2O(l)$$

12.4 State symbols and other information

The state (phase) of the compounds can be expressed in the chemical equation. This is done by placing the correct label on the right hand side of the formula. There are only four labels that can be used:

1. (g) for gaseous compounds
2. (l) for liquids
3. (s) for solid compounds
4. (aq) for an aqueous (water) solution

Occasionally, a catalyst is added to the reaction. A catalyst is a substance that speeds up the reaction without undergoing any change to itself. In a chemical equation, this is shown by using the symbol of the catalyst above the arrow in the equation.

To show that heat was needed for the reaction, a Greek delta (Δ) is placed above the arrow in the same way as the catalyst.

Important: You may remember from chapter 11 that energy cannot be created or destroyed during a chemical reaction but it may change form. In an exothermic reaction, ΔH is less than zero, and in an endothermic reaction, ΔH is greater than zero. This value is often written at the end of a chemical equation.

Worked Example 53: Balancing chemical equations 4

Question: Solid zinc metal reacts with aqueous hydrochloric acid to form an aqueous solution of zinc chloride ($ZnCl_2$)and hydrogen gas. Write a balanced equation for this reaction.

Answer

Step 1 : Identify the reactants and products and their chemical formulae
The reactants are zinc (Zn) and hydrochloric acid (HCl). The products are zinc chloride ($ZnCl_2$) and hydrogen (H_2).

Step 2 : Place the reactants on the left of the equation and the products on the right hand side of the arrow.

$$Zn + HCl \rightarrow ZnCl_2 + H_2$$

Step 3 : Balance the equation
You will notice that the zinc atoms balance but the chlorine and hydrogen atoms don't. Since there are two chlorine atoms on the right and only one on the left, we will give HCl a coefficient of 2 so that there will be two chlorine atoms on each side of the equation.

$$Zn + 2HCl \rightarrow ZnCl_2 + H_2$$

Step 4 : Check that all the atoms balance
When you look at the equation again, you will see that all the atoms are now balanced.

Step 5 : Ensure all details (e.g. state symbols) are added
In the initial description, you were told that zinc was a metal, hydrochloric acid and zinc chloride were in aqueous solutions and hydrogen was a gas.

$$Zn(s) + 2HCl(aq) \rightarrow ZnCl_2(aq) + H_2(g)$$

Worked Example 54: Balancing chemical equations 5 (advanced)

Question: Balance the following equation:

$$(NH_4)_2SO_4 + NaOH \rightarrow NH_3 + H_2O + Na_2SO_4$$

In this example, the first two steps are not necessary because the reactants and products have already been given.

Answer

Step 1 : Balance the equation

With a complex equation, it is always best to start with atoms that appear only once on each side i.e. Na, N and S atoms. Since the S atoms already balance, we will start with Na and N atoms. There are two Na atoms on the right and one on the left. We will add a second Na atom by giving NaOH a coefficient of two. There are two N atoms on the left and one on the right. To balance the N atoms, NH_3 will be given a coefficient of two. The equation now looks as follows:

$$(NH_4)_2SO_4 + 2NaOH \rightarrow 2NH_3 + H_2O + Na_2SO_4$$

Step 2 : Check that all atoms balance

N, Na and S atoms balance, but O and H atoms do not. There are six O atoms and ten H atoms on the left, and five O atoms and eight H atoms on the right. We need to add one O atom and two H atoms on the right to balance the equation. This is done by adding another H_2O molecule on the right hand side. We now need to check the equation again:

$$(NH_4)_2SO_4 + 2NaOH \rightarrow 2NH_3 + 2H_2O + Na_2SO_4$$

The equation is now balanced.

Exercise: Balancing more advanced chemical equations

Write balanced equations for each of the following reactions:

1. $Al_2O_3(s) + H_2SO_4(aq) \rightarrow Al_2(SO_4)_3(aq) + 3H_2O(l)$
2. $Mg(OH)_2(aq) + HNO_3(aq) \rightarrow Mg(NO_3)_2(aq) + 2H_2O(l)$
3. Lead(II)nitrate solution reacts with potassium iodide solution.
4. When heated, aluminium reacts with solid copper oxide to produce copper metal and aluminium oxide (Al_2O_3).
5. When calcium chloride solution is mixed with silver nitrate solution, a white precipitate (solid) of silver chloride appears. Calcium nitrate ($Ca(NO_3)_2$) is also produced in the solution.

12.5 Summary

- A **chemical equation** uses symbols to describe a chemical reaction.
- In a chemical equation, **reactants** are written on the left hand side of the equation, and the **products** on the right. The arrow is used to show the direction of the reaction.
- When representing chemical change, it is important to be able to write the **chemical formula** of a compound.
- In any chemical reaction, the **law of conservation of mass** applies. This means that the total atomic mass of the reactants must be the same as the total atomic mass of the products. This also means that the number of atoms of each element in the reactants must be the same as the number of atoms of each element in the product.
- If the number of atoms of each element in the reactants is the same as the number of atoms of each element in the product, then the equation is **balanced**.
- If the number of atoms of each element in the reactants is not the same as the number of atoms of each element in the product, then the equation is **not balanced**.
- In order to balance an equation, **coefficients** can be placed in front of the reactants and products until the number of atoms of each element is the same on both sides of the equation.

Exercise: Summary exercise

Balance each of the following chemical equations:

1. $NH_4 + H_2O \rightarrow NH_4OH$
2. Sodium chloride and water react to form sodium hydroxide, chlorine and hydrogen.
3. Propane is a fuel that is commonly used as a heat source for engines and homes. Balance the following equation for the combustion of propane:
$$C_3H_8(l) + O_2(g) \rightarrow CO_2(g) + H_2O(l)$$
4. Aspartame, an artificial sweetener, has the formula $C_{14}H_{18}N_2O_5$. Write the balanced equation for its combustion (reaction with O_2) to form CO_2 gas, liquid H_2O, and N_2 gas.
5. $Fe_2(SO_4)_3 + K(SCN) \rightarrow K_3Fe(SCN)_6 + K_2SO_4$
6. Chemical weapons were banned by the Geneva Protocol in 1925. According to this protocol, all chemicals that release suffocating and poisonous gases are not to be used as weapons. White phosphorus, a very reactive allotrope of phosphorus, was recently used during a military attack. Phosphorus burns vigorously in oxygen. Many people got severe burns and some died as a result. The equation for this spontaneous reaction is:
$$P_4(s) + O_2(g) \rightarrow P_2O_5(s)$$
   (a) Balance the chemical equation.
   (b) Prove that the law of conservation of mass is obeyed during this chemical reaction.
   (c) Name the product formed during this reaction.
   (d) Classify the reaction as endothermic or exothermic. Give a reason for your answer.
   (e) Classify the reaction as a sythesis or decomposition reaction. Give a reason for your answer.

   (DoE Exemplar Paper 2 2007)

Chapter 13

Quantitative Aspects of Chemical Change - Grade 11

An equation for a chemical reaction can provide us with a lot of useful information. It tells us what the reactants and the products are in the reaction, and it also tells us the ratio in which the reactants combine to form products. Look at the equation below:

$$Fe + S \rightarrow FeS$$

In this reaction, every atom of iron (Fe) will react with a single atom of sulfur (S) to form one molecule of iron sulfide (FeS). However, what the equation doesn't tell us, is the **quantities** or the **amount** of each substance that is involved. You may for example be given a small sample of iron for the reaction. How will you know how many atoms of iron are in this sample? And how many atoms of sulfur will you need for the reaction to use up all the iron you have? Is there a way of knowing what mass of iron sulfide will be produced at the end of the reaction? These are all very important questions, especially when the reaction is an industrial one, where it is important to know the quantities of reactants that are needed, and the quantity of product that will be formed. This chapter will look at how to quantify the changes that take place in chemical reactions.

13.1 The Mole

Sometimes it is important to know exactly how many particles (e.g. atoms or molecules) are in a sample of a substance, or what quantity of a substance is needed for a chemical reaction to take place.

You will remember from chapter 3 that the **relative atomic mass** of an element, describes the mass of an atom of that element relative to the mass of an atom of carbon-12. So the mass of an atom of carbon (relative atomic mass is 12 u) for example, is twelve times greater than the mass of an atom of hydrogen, which has a relative atomic mass of 1 u. How can this information be used to help us to know what mass of each element will be needed if we want to end up with the same number of *atoms* of carbon and hydrogen?

Let's say for example, that we have a sample of 12g carbon. What mass of *hydrogen* will contain the same number of atoms as 12 g carbon? We know that each atom of carbon weighs twelve times more than an atom of hydrogen. Surely then, we will only need 1g of hydrogen for the number of atoms in the two samples to be the same? You will notice that the number of particles (in this case, *atoms*) in the two substances is the same when the ratio of their sample masses (12g carbon: 1g hydrogen = 12:1) is the same as the ratio of their relative atomic masses (12 u: 1 u = 12:1).

To take this a step further, if you were to weigh out samples of a number of elements so that the mass of the sample was the same as the relative atomic mass of that element, you would find that the number of particles in each sample is 6.023 x 10^{23}. These results are shown in table 13.1 below for a number of different elements. So, 24.31 g of magnesium (relative atomic mass = 24.31 u) for example, has the same number of atoms as 40.08 g of calcium (relative atomic mass = 40.08 u).

Table 13.1: Table showing the relationship between the sample mass, the relative atomic mass and the number of atoms in a sample, for a number of elements.

Element	Relative atomic mass (u)	Sample mass (g)	Atoms in sample
Hydrogen (H)	1.01	1.01	6.023 x 10^{23}
Carbon (C)	12.01	12.01	6.023 x 10^{23}
Magnesium (Mg)	24.31	24.31	6.023 x 10^{23}
Sulfur (S)	32.07	32.07	6.023 x 10^{23}
Calcium (Ca)	40.08	40.08	6.023 x 10^{23}

This result is so important that scientists decided to use a special unit of measurement to define this quantity: the **mole** or 'mol'. A **mole** is defined as being an amount of a substance which contains the same number of particles as there are atoms in 12 g of carbon. In the examples that were used earlier, 24.31 g magnesium is *one mole* of magnesium, while 40.08 g of calcium is *one mole* of calcium. A mole of any substance always contains the same number of particles.

Definition: Mole
The mole (abbreviation 'n') is the SI (Standard International) unit for 'amount of substance'. It is defined as an amount of substance that contains the same number of particles (atoms, molecules or other particle units) as there are atoms in 12 g carbon.

In one mole of any substance, there are 6.023 x 10^{23} particles. This is known as **Avogadro's number**.

Definition: Avogadro constant
The number of particles in a mole, equal to 6.023 x 10^{23}. It is also sometimes referred to as the number of atoms in 12 g of carbon-12.

Interesting Fact

The original hypothesis that was proposed by Amadeo Avogadro was that *'equal volumes of gases, at the same temperature and pressure, contain the same number of molecules'*. His ideas were not accepted by the scientific community and it was only four years after his death, that his original hypothesis was accepted and that it became known as 'Avogadro's Law'. In honour of his contribution to science, the number of particles in one mole was named *Avogadro's number*.

Exercise: Moles and mass

1. Complete the following table:

Element	Relative atomic mass (u)	Sample mass (g)	Number of moles in the sample
Hydrogen	1.01	1.01	
Magnesium	24.31	24.31	
Carbon	12.01	24.02	
Chlorine	35.45	70.9	
Nitrogen		42.08	

2. How many atoms are there in...
   (a) 1 mole of a substance
   (b) 2 moles of calcium
   (c) 5 moles of phosphorus
   (d) 24.31 g of magnesium
   (e) 24.02 g of carbon

13.2 Molar Mass

Definition: Molar mass
Molar mass (M) is the mass of 1 mole of a chemical substance. The unit for molar mass is **grams per mole** or $g.mol^{-1}$.

Refer to table 13.1. You will remember that when the mass, in grams, of an element is equal to its relative atomic mass, the sample contains one mole of that element. This mass is called the **molar mass** of that element.

It is worth remembering the following: On the Periodic Table, the relative atomic mass that is shown can be interpreted in two ways.

1. The mass of a *single, average atom* of that element relative to the mass of an atom of carbon.

2. The mass of *one mole of the element*. This second use is the molar mass of the element.

Table 13.2: The relationship between relative atomic mass, molar mass and the mass of one mole for a number of elements.

Element	Relative atomic mass (u)	Molar mass ($g.mol^{-1}$)	Mass of one mole of the element (g)
Magnesium	24.31	24.31	24.31
Lithium	6.94	6.94	6.94
Oxygen	16	16	16
Nitrogen	14.01	14.01	14.01
Iron	55.85	55.85	55.85

Worked Example 55: Calculating the number of moles from mass

Question: Calculate the number of moles of iron (Fe) in a 111.7 g sample.

Answer
Step 1 : Find the molar mass of iron
If we look at the periodic table, we see that the molar mass of iron is 55.85 g.mol^{-1}. This means that 1 mole of iron will have a mass of 55.85 g.

Step 2 : Use the molar mass and sample mass to calculate the number of moles of iron
If 1 mole of iron has a mass of 55.85 g, then: the number of moles of iron in 111.7 g must be:

$$\frac{111.7g}{55.85g.mol^{-1}} = 2mol$$

There are 2 moles of iron in the sample.

Worked Example 56: Calculating mass from moles

Question: You have a sample that contains 5 moles of zinc.

1. What is the mass of the zinc in the sample?
2. How many atoms of zinc are in the sample?

Answer
Step 1 : Find the molar mass of zinc
Molar mass of zinc is 65.38 g.mol^{-1}, meaning that 1 mole of zinc has a mass of 65.38 g.

Step 2 : Calculate the mass of zinc, using moles and molar mass.
If 1 mole of zinc has a mass of 65.38 g, then 5 moles of zinc has a mass of:

65.38 g x 5 mol = 326.9 g (answer to a)

Step 3 : Use the number of moles of zinc and Avogadro's number to calculate the number of zinc atoms in the sample.

$$5 \times 6.023 \times 10^{23} = 30.115 \times 10^{23}$$

Exercise: Moles and molar mass

1. Give the molar mass of each of the following elements:
   (a) hydrogen
   (b) nitrogen
   (c) bromine
2. Calculate the number of moles in each of the following samples:
   (a) 21.62 g of boron (B)

(b) 54.94 g of manganese (Mn)
(c) 100.3 g of mercury (Hg)
(d) 50 g of barium (Ba)
(e) 40 g of lead (Pb)

13.3 An equation to calculate moles and mass in chemical reactions

The calculations that have been used so far, can be made much simpler by using the following equation:

$$\textbf{n} \text{ (number of moles)} = \frac{\textbf{m} \text{ (mass of substance in g)}}{\textbf{M} \text{ (molar mass of substance in } \mathrm{g \cdot mol^{-1}})}$$

Important: Remember that when you use the equation n = m/M, the mass is always in grams (g) and molar mass is in grams per mol (g.mol^{-1}).

The equation can also be used to calculate mass and molar mass, using the following equations:

$$m = n \times M$$

and

$$M = \frac{m}{n}$$

The following diagram may help to remember the relationship between these three variables. You need to imagine that the horizontal line is like a 'division' sign and that the vertical line is like a 'multiplication' sign. So, for example, if you want to calculate 'M', then the remaining two letters in the triangle are 'm' and 'n' and 'm' is above 'n' with a division sign between them. In your calculation then, 'm' will be the numerator and 'n' will be the denominator.

Worked Example 57: Calculating moles from mass

Question: Calculate the number of moles of copper there are in a sample that weighs 127 g.

Answer
Step 1 : Write the equation to calculate the number of moles

$$n = \frac{m}{M}$$

Step 2 : Substitute numbers into the equation

$$n = \frac{127}{63.55} = 2$$

There are 2 moles of copper in the sample.

Worked Example 58: Calculating mass from moles

Question: You are given a 5 mol sample of sodium. What mass of sodium is in the sample?

Answer
Step 1 : Write the equation to calculate the sample mass.

$$m = n \times M$$

Step 2 : Substitute values into the equation.
M_{Na} = 22.99 g.mol^{-1}

Therefore,

$$m = 5 \times 22.99 = 114.95g$$

The sample of sodium has a mass of 114.95 g.

Worked Example 59: Calculating atoms from mass

Question: Calculate the number of atoms there are in a sample of aluminium that weighs 80.94 g.

Answer
Step 1 : Calculate the number of moles of aluminium in the sample.

$$n = \frac{m}{M} = \frac{80.94}{26.98} = 3moles$$

Step 2 : Use Avogadro's number to calculate the number of atoms in the sample.
Number of atoms in 3 mol aluminium = 3 × 6.023 × 10^{23}
There are 18.069 × 10^{23} aluminium atoms in a sample of 80.94 g.

Exercise: Some simple calculations

1. Calculate the number of moles in each of the following samples:
   (a) 5.6 g of calcium
   (b) 0.02 g of manganese
   (c) 40 g of aluminium
2. A lead sinker has a mass of 5 g.
   (a) Calculate the number of moles of lead the sinker contains.
   (b) How many lead atoms are in the sinker?
3. Calculate the mass of each of the following samples:
   (a) 2.5 mol magnesium
   (b) 12 g lithium
   (c) 4.5 $\times$ 10^{25} atoms of silica

13.4 Molecules and compounds

So far, we have only discussed moles, mass and molar mass in relation to *elements*. But what happens if we are dealing with a molecule or some other chemical compound? Do the same concepts and rules apply? The answer is 'yes'. However, you need to remember that all your calculations will apply to the *whole molecule*. So, when you calculate the molar mass of a molecule, you will need to add the molar mass of each atom in that compound. Also, the number of moles will also apply to the whole molecule. For example, if you have one mole of nitric acid (HNO_3), it means you have 6.023 x 10^{23} **molecules** of nitric acid in the sample. This also means that there are 6.023 $\times$ 10^{23} **atoms** of hydrogen, 6.023 $\times$ 10^{23} **atoms** of nitrogen and (3 $\times$ 6.023 $\times$ 10^{23}) **atoms** of oxygen in the sample.

In a balanced chemical equation, the number that is written in front of the element or compound, shows the **mole ratio** in which the reactants combine to form a product. If there are no numbers in front of the element symbol, this means the number is '1'.

$$\text{e.g. } N_2 + 3H_2 \rightarrow 2NH_3$$

In this reaction, 1 mole of nitrogen reacts with 3 moles of hydrogen to produce 2 moles of ammonia.

Worked Example 60: Calculating molar mass

Question: Calculate the molar mass of H_2SO_4.

Answer
Step 1 : Use the periodic table to find the molar mass for each element in the molecule.
Hydrogen = 1.008 g.mol^{-1}; Sulfur = 32.07 g.mol^{-1}; Oxygen = 16 g.mol^{-1}

Step 2 : Add the molar masses of each atom in the molecule

$$M_{(H_2SO_4)} = (2 \times 1.008) + (32.07) + (4 \times 16) = 98.09 g.mol^{-1}$$

Worked Example 61: Calculating moles from mass

Question: Calculate the number of moles there are in 1kg of $MgCl_2$.

Answer
Step 1 : Write the equation for calculating the number of moles in the sample.

$$n = \frac{m}{M}$$

Step 2 : Calculate the values that you will need, to substitute into the equation

1. Convert mass into grams

$$m = 1kg \times 1000 = 1000g$$

2. Calculate the molar mass of $MgCl_2$.

$$M_{(MgCl_2)} = 24.31 + (2 \times 35.45) = 95.21 g.mol^{-1}$$

Step 3 : Substitute values into the equation

$$n = \frac{1000}{95.21} = 10.5 mol$$

There are 10.5 moles of magnesium chloride in a 1 kg sample.

Worked Example 62: Calculating the mass of reactants and products

Question: Barium chloride and sulfuric acid react according to the following equation to produce barium sulphate and hydrochloric acid.

$$BaCl_2 + H_2SO_4 \rightarrow BaSO_4 + 2HCl$$

If you have 2 g of $BaCl_2$...

1. What quantity (in g) of H_2SO_4 will you need for the reaction so that all the barium chloride is used up?
2. What mass of HCl is produced during the reaction?

Answer
Step 1 : Calculate the number of moles of $BaCl_2$ that react.

$$n = \frac{m}{M} = \frac{2}{208.24} = 0.0096 mol$$

Step 2 : Determine how many moles of H_2SO_4 are needed for the reaction
According to the balanced equation, 1 mole of $BaCl_2$ will react with 1 mole of H_2SO_4. Therefore, if 0.0096 moles of $BaCl_2$ react, then there must be the same number of moles of H_2SO_4 that react because their mole ratio is 1:1.

Step 3 : Calculate the mass of H_2SO_4 that is needed.

$$m = n \times M = 0.0096 \times 98.086 = 0.94g$$

(answer to 1)

Step 4 : Determine the number of moles of HCl produced.
According to the balanced equation, 2 moles of HCl are produced for every 1 mole of the two reactants. Therefore the number of moles of HCl produced is (2 × 0.0096), which equals 0.0192 moles.

Step 5 : Calculate the mass of HCl.

$$m = n \times M = 0.0192 \times 35.73 = 0.69g$$

(answer to 2)

Activity :: Group work : Understanding moles, molecules and Avogadro's number

Divide into groups of three and spend about 20 minutes answering the following questions together:

1. What are the units of the mole? Hint: Check the definition of the mole.
2. You have a 56 g sample of iron sulfide (FeS)
   (a) How many **moles** of FeS are there in the sample?
   (b) How many **molecules** of FeS are there in the sample?
   (c) What is the difference between a mole and a molecule?
3. The exact size of **Avogadro's number** is sometimes difficult to imagine.
   (a) Write down Avogadro's number without using scientific notation.
   (b) How long would it take to count to Avogadro's number? You can assume that you can count two numbers in each second.

?

Exercise: More advanced calculations

1. Calculate the molar mass of the following chemical compounds:
   (a) KOH
   (b) $FeCl_3$
   (c) $Mg(OH)_2$
2. How many moles are present in:
   (a) 10 g of Na_2SO_4
   (b) 34 g of $Ca(OH)_2$
   (c) 2.45 x 10^{23} molecules of CH_4?

3. For a sample of 0.2 moles of potassium bromide (KBr), calculate...
   (a) the number of moles of K^+ ions
   (b) the number of moles of Br^- ions
4. You have a sample containing 3 moles of calcium chloride.
   (a) What is the chemical formula of calcium chloride?
   (b) How many calcium atoms are in the sample?
5. Calculate the mass of:
   (a) 3 moles of NH_4OH
   (b) 4.2 moles of $Ca(NO_3)_2$
6. 96.2 g sulfur reacts with an unknown quantity of zinc according to the following equation:

$$Zn + S \rightarrow ZnS$$

   (a) What mass of zinc will you need for the reaction, if all the sulfur is to be used up?
   (b) What mass of zinc sulfide will this reaction produce?
7. Calcium chloride reacts with carbonic acid to produce calcium carbonate and hydrochloric acid according to the following equation:

$$CaCl_2 + H_2CO_3 \rightarrow CaCO_3 + 2HCl$$

   If you want to produce 10 g of calcium carbonate through this chemical reaction, what quantity (in g) of calcium chloride will you need at the start of the reaction?

13.5 The Composition of Substances

The **empirical formula** of a chemical compound is a simple expression of the relative number of each type of atom in it. In contrast, the **molecular formula** of a chemical compound gives the actual number of atoms of each element found in a molecule of that compound.

Definition: Empirical formula
The empirical formula of a chemical compound gives the relative number of each type of atom in it.

Definition: Molecular formula
The molecular formula of a chemical compound gives the exact number of atoms of each element in one molecule of that compound.

The compound *ethanoic acid* for example, has the molecular formula CH_3COOH or simply $C_2H_4O_2$. In one molecule of this acid, there are two carbon atoms, four hydrogen atoms and two oxygen atoms. The ratio of atoms in the compound is 2:4:2, which can be simplified to 1:2:1. Therefore, the empirical formula for this compound is CH_2O. The empirical formula contains the smallest whole number ratio of the elements that make up a compound.

Knowing either the empirical or molecular formula of a compound, can help to determine its composition in more detail. The opposite is also true. Knowing the *composition* of a substance can help you to determine its formula. There are three different types of composition problems that you might come across:

1. Problems where you will be given the formula of the substance and asked to calculate the percentage by mass of each element in the substance.

2. Problems where you will be given the percentage composition and asked to calculate the formula.

3. Problems where you will be given the products of a chemical reaction and asked to calculate the formula of one of the reactants. These are usually referred to as combustion analysis problems.

Worked Example 63: Calculating the percentage by mass of elements in a compound

Question: Calculate the percentage that each element contributes to the overall mass of sulfuric acid (H_2SO_4).

Answer

Step 1 : Write down the relative atomic mass of each element in the compound.

Hydrogen = 1.008 × 2 = 2.016 u
Sulfur = 32.07 u
Oxygen = 4 × 16 = 64 u

Step 2 : Calculate the molecular mass of sulfuric acid.

Use the calculations in the previous step to calculate the molecular mass of sulfuric acid.

$$Mass = 2.016 + 32.07 + 64 = 98.09u$$

Step 3 : Convert the mass of each element to a percentage of the total mass of the compound

Use the equation:

Percentage by mass = atomic mass / molecular mass of H_2SO_4 × 100%

Hydrogen

$$\frac{2.016}{98.09} \times 100\% = 2.06\%$$

Sulfur

$$\frac{32.07}{98.09} \times 100\% = 32.69\%$$

Oxygen

$$\frac{64}{98.09} \times 100\% = 65.25\%$$

(You should check at the end that these percentages add up to 100%!)
In other words, in one molecule of sulfuric acid, hydrogen makes up 2.06% of the mass of the compound, sulfur makes up 32.69% and oxygen makes up 65.25%.

Worked Example 64: Determining the empirical formula of a compound

Question: A compound contains 52.2% carbon (C), 13.0% hydrogen (H) and 34.8% oxygen (O). Determine its empirical formula.

Answer
Step 1 : If we assume that we have 100 g of this substance, then we can convert each element percentage into a mass in grams.
Carbon = 52.2 g, hydrogen = 13 g and oxygen = 34.8 g

Step 2 : Convert the mass of each element into number of moles

$$n = \frac{m}{M}$$

Therefore,

$$n(carbon) = \frac{52.2}{12.01} = 4.35mol$$

$$n(hydrogen) = \frac{13}{1.008} = 12.90mol$$

$$n(oxygen) = \frac{34.8}{16} = 2.18mol$$

Step 3 : Convert these numbers to the simplest mole ratio by dividing by the smallest number of moles
In this case, the smallest number of moles is 2.18. Therefore...
Carbon

$$\frac{4.35}{2.18} = 2$$

Hydrogen

$$\frac{12.90}{2.18} = 6$$

Oxygen

$$\frac{2.18}{2.18} = 1$$

Therefore the empirical formula of this substance is: C_2H_6O. Do you recognise this compound?

Worked Example 65: Determining the formula of a compound

Question: 207 g of lead combines with oxygen to form 239 g of a lead oxide. Use this information to work out the formula of the lead oxide (Relative atomic masses: Pb = 207 u and O = 16 u).

Answer
Step 1 : Calculate the mass of oxygen in the reactants

$$239 - 207 = 32g$$

Step 2 : Calculate the number of moles of lead and oxygen in the reactants.

$$n = \frac{m}{M}$$

Lead

$$\frac{207}{207} = 1mol$$

Oxygen

$$\frac{32}{16} = 2mol$$

Step 3 : Deduce the formula of the compound
The mole ratio of Pb:O in the product is 1:2, which means that for every atom of lead, there will be two atoms of oxygen. The formula of the compound is PbO_2.

Worked Example 66: Empirical and molecular formula

Question: Vinegar, which is used in our homes, is a dilute form of acetic acid. A sample of acetic acid has the following percentage composition: 39.9% carbon, 6.7% hyrogen and 53.4% oxygen.

1. Determine the empirical formula of acetic acid.
2. Determine the molecular formula of acetic acid if the molar mass of acetic acid is 60g/mol.

Answer
Step 1 : Calculate the mass of each element in 100 g of acetic acid.
In 100g of acetic acid, there is 39.9 g C, 6.7 g H and 53.4 g O

Step 2 : Calculate the number of moles of each element in 100 g of acetic acid.
$n = \frac{m}{M}$

$$\begin{aligned} n_C &= \frac{39.9}{12} = 3.33\,\text{mol} \\ n_H &= \frac{6.7}{1} = 6.7\,\text{mol} \\ n_O &= \frac{53.4}{16} = 3.34\,\text{mol} \end{aligned}$$

Step 3 : Divide the number of moles of each element by the lowest number to get the simplest mole ratio of the elements (i.e. the empirical formula) in acetic acid.
Empirical formula is CH_2O

Step 4 : Calculate the molecular formula, using the molar mass of acetic acid.
The molar mass of acetic acid using the empirical formula is 30 g/mol. Therefore the actual number of moles of each element must be double what it is in the emprical formula.

The molecular formula is therefore $C_2H_4O_2$ or CH_3COOH

?

Exercise: Moles and empirical formulae

1. Calcium chloride is produced as the product of a chemical reaction.
   (a) What is the formula of calcium chloride?
   (b) What percentage does each of the elements contribute to the mass of a molecule of calcium chloride?
   (c) If the sample contains 5 g of calcium chloride, what is the mass of calcium in the sample?
   (d) How many moles of calcium chloride are in the sample?
2. 13g of zinc combines with 6.4g of sulfur.What is the empirical formula of zinc sulfide?
   (a) What mass of zinc sulfide will be produced?
   (b) What percentage does each of the elements in zinc sulfide contribute to its mass?
   (c) Determine the formula of zinc sulfide.
3. A calcium mineral consisted of 29.4% calcium, 23.5% sulphur and 47.1% oxygen by mass. Calculate the empirical formula of the mineral.
4. A chlorinated hydrocarbon compound when analysed, consisted of 24.24% carbon, 4.04% hydrogen, 71.72% chlorine. The molecular mass was found to be 99 from another experiment. Deduce the empirical and molecular formula.

13.6 Molar Volumes of Gases

It is possible to calculate the volume of a mole of gas at STP using what we now know about gases.

1. *Write down the ideal gas equation*

   pV = nRT, therefore V = $\frac{nRT}{p}$

2. *Record the values that you know, making sure that they are in SI units*

   You know that the gas is under STP conditions. These are as follows:

   p = 101.3 kPa = 101300 Pa

   n = 1 mole

   R = 8.3 J.K^{-1}.mol^{-1}

   T = 273 K

3. *Substitute these values into the original equation.*

$$V = \frac{nRT}{p}$$

$$V = \frac{1mol \times 8.3J.K^{-1}.mol^{-1} \times 273K}{101300Pa}$$

4. *Calculate the volume of 1 mole of gas under these conditions*

   The volume of 1 mole of gas at STP is 22.4 $\times$ 10^{-3} m^3 = 22.4 dm^3.

Important: The standard units used for this equation are P in Pa, V in m^3 and T in K. Remember also that 1000cm^3 = 1dm^3 and 1000dm^3 = 1m^3.

Worked Example 67: Ideal Gas

Question: A sample of gas occupies a volume of 20 dm^3, has a temperature of 280 K and has a pressure of 105 Pa. Calculate the number of moles of gas that are present in the sample.
Answer
Step 1 : Convert all values into SI units
The only value that is not in SI units is volume. V = 0.02 m^3.

Step 2 : Write the equation for calculating the number of moles in a gas.
We know that pV = nRT

Therefore,

$$n = \frac{pV}{RT}$$

Step 3 : Substitute values into the equation to calculate the number of moles of the gas.

$$n = \frac{105 \times 0.02}{8.31 \times 280} = \frac{2.1}{2326.8} = 0.0009\text{moles}$$

Exercise: Using the combined gas law

1. An enclosed gas has a volume of 300 cm^3 and a temperature of 300 K. The pressure of the gas is 50 kPa. Calculate the number of moles of gas that are present in the container.
2. What pressure will 3 mol gaseous nitrogen exert if it is pumped into a container that has a volume of 25 dm^3 at a temperature of 29 ^{0}C?
3. The volume of air inside a tyre is 19 litres and the temperature is 290 K. You check the pressure of your tyres and find that the pressure is 190 kPa. How many moles of air are present in the tyre?
4. Compressed carbon dioxide is contained within a gas cylinder at a pressure of 700 kPa. The temperature of the gas in the cylinder is 310 K and the number of moles of gas is 13 moles carbon dioxide. What is the volume of the gas inside?

13.7 Molar concentrations in liquids

A typical solution is made by dissolving some solid substance in a liquid. The amount of substance that is dissolved in a given volume of liquid is known as the **concentration** of the liquid. Mathematically, concentration (C) is defined as moles of solute (n) per unit volume (V) of solution.

$$C = \frac{n}{V}$$

For this equation, the units for volume are dm^3. Therefore, the unit of concentration is mol.dm^{-3}. When concentration is expressed in mol.dm^{-3} it is known as the **molarity** (M) of the solution. Molarity is the most common expression for concentration.

Definition: Concentration
Concentration is a measure of the amount of solute that is dissolved in a given volume of liquid. It is measured in mol.dm^{-3}. Another term that is used for concentration is **molarity (M)**

Worked Example 68: Concentration Calculations 1

Question: If 3.5 g of sodium hydroxide (NaOH) is dissolved in 2.5 dm^3 of water, what is the concentration of the solution in mol.dm^{-3}?

Answer
Step 1 : Convert the mass of NaOH into moles

$$n = \frac{m}{M} = \frac{3.5}{40} = 0.0875mol$$

Step 2 : Calculate the concentration of the solution.

$$C = \frac{n}{V} = \frac{0.0875}{2.5} = 0.035$$

The concentration of the solution is 0.035 mol.dm^{-3} or 0.035 M

Worked Example 69: Concentration Calculations 2

Question: You have a 1 dm^3 container in which to prepare a solution of potassium permanganate ($KMnO_4$). What mass of $KMnO_4$ is needed to make a solution with a concentration of 0.2 M?

Answer
Step 1 : Calculate the number of moles of $KMnO_4$ needed.

$$C = \frac{n}{V}$$

therefore

$$n = C \times V = 0.2 \times 1 = 0.2mol$$

Step 2 : Convert the number of moles of $KMnO_4$ to mass.

$$m = n \times M = 0.2 \times 158.04 = 31.61g$$

The mass of $KMnO_4$ that is needed is 31.61 g.

Worked Example 70: Concentration Calculations 3

Question: How much sodium chloride (in g) will one need to prepare 500 cm^3 of solution with a concentration of 0.01 M?

Answer
Step 1 : Convert all quantities into the correct units for this equation.

$$V = \frac{500}{1000} = 0.5dm^3$$

Step 2 : Calculate the number of moles of sodium chloride needed.

$$n = C \times V = 0.01 \times 0.5 = 0.005mol$$

Step 3 : Convert moles of $KMnO_4$ to mass.

$$m = n \times M = 0.005 \times 58.45 = 0.29g$$

The mass of sodium chloride needed is 0.29 g

Exercise: Molarity and the concentration of solutions

1. 5.95g of potassium bromide was dissolved in 400cm3 of water. Calculate its molarity.
2. 100 g of sodium chloride (NaCl) is dissolved in 450 cm^3 of water.
   (a) How many moles of NaCl are present in solution?
   (b) What is the volume of water (in dm^3)?
   (c) Calculate the concentration of the solution.
   (d) What mass of sodium chloride would need to be added for the concentration to become 5.7 $mol.dm^{-3}$?
3. What is the molarity of the solution formed by dissolving 80 g of sodium hydroxide (NaOH) in 500 cm^3 of water?
4. What mass (g) of hydrogen chloride (HCl) is needed to make up 1000 cm^3 of a solution of concentration 1 $mol.dm^{-3}$?
5. How many moles of H_2SO_4 are there in 250 cm^3 of a 0.8M sulphuric acid solution? What mass of acid is in this solution?

13.8 Stoichiometric calculations

Stoichiometry is the study and calculation of relationships between reactants and products of chemical reactions. Chapter 12 showed how to write balanced chemical equations. By knowing the ratios of substances in a reaction, it is possible to use stoichiometry to calculate the amount of reactants and products that are involved in the reaction. Some examples are shown below.

Worked Example 71: Stoichiometric calculation 1

Question: What volume of oxygen at S.T.P. is needed for the complete combustion of 2dm^3 of propane (C_3H_8)? (Hint: CO_2 and H_2O are the products in this reaction)

Answer
Step 1 : Write a balanced equation for the reaction.

$$C_3H_8(g) + 5O_2(g) \rightarrow 3CO_2(g) + 4H_2O(g)$$

Step 2 : Determine the ratio of oxygen to propane that is needed for the reaction.
From the balanced equation, the ratio of oxygen to propane in the reactants is 5:1.

Step 3 : Determine the volume of oxygen needed for the reaction.
1 volume of propane needs 5 volumes of oxygen, therefore 2 dm^3 of propane will need 10 dm^3 of oxygen for the reaction to proceed to completion.

Worked Example 72: Stoichiometric calculation 2

Question: What mass of iron (II) sulphide is formed when 5.6 g of iron is completely reacted with sulfur?

Answer
Step 1 : Write a balanced chemical equation for the reaction.

$$Fe(s) + S(s) \rightarrow FeS(s)$$

Step 2 : Calculate the number of moles of iron that react.

$$n = \frac{m}{M} = \frac{5.6}{55.85} = 0.1mol$$

Step 3 : Determine the number of moles of FeS produced.
From the equation 1 mole of Fe gives 1 mole of FeS. Therefore, 0.1 moles of iron in the reactants will give 0.1 moles of iron sulfide in the product.

Step 4 : Calculate the mass of iron sulfide formed

$$m = n \times M = 0.1 \times 87.911 = 8.79g$$

The mass of iron (II) sulfide that is produced during this reaction is 8.79 g.

Important:
A closer look at the previous worked example shows that 5.6 g of iron is needed to produce 8.79 g of iron (II) sulphide. The amount of sulfur that is needed in the reactants is 3.2 g. What would happen if the amount of sulfur in the reactants was increased to 6.4 g but the amount of iron was still 5.6 g? Would more FeS be produced? In fact, the amount of iron(II) sulfide produced remains the same. No matter how much sulfur is added to the system, the amount of iron (II) sulfide will not increase because there is not enough iron to react with the additional sulfur in the reactants to produce more FeS. When all the iron is used up the reaction stops. In this example, the iron is called the **limiting reagent**. Because there is more sulfur than can be used up in the reaction, it is called the **excess reagent**.

Worked Example 73: Industrial reaction to produce fertiliser

Question: Sulfuric acid (H_2SO_4) reacts with ammonia (NH_3) to produce the fertiliser ammonium sulphate ($(NH_4)_2SO_4$) according to the following equation:

$$H_2SO_4(aq) + 2NH_3(g) \rightarrow (NH_4)_2SO_4(aq)$$

What is the maximum mass of ammonium sulphate that can be obtained from 2.0 kg of sulfuric acid and 1.0 kg of ammonia?

Answer

Step 1 : Convert the mass of sulfuric acid and ammonia into moles

$$n(H_2SO_4) = \frac{m}{M} = \frac{2000g}{98.078g/mol} = 20.39mol$$

$$n(NH_3) = \frac{1000g}{17.03g/mol} = 58.72mol$$

Step 2 : Use the balanced equation to determine which of the reactants is limiting.

From the balanced chemical equation, 1 mole of H_2SO_4 reacts with 2 moles of NH_3 to give 1 mole of $(NH_4)_2SO_4$. Therefore 20.39 moles of H_2SO_4 need to react with 40.78 moles of NH_3. In this example, NH_3 is in excess and H_2SO_4 is the limiting reagent.

Step 3 : Calculate the maximum amount of ammonium sulphate that can be produced

Again from the equation, the mole ratio of H_2SO_4 in the reactants to $(NH_4)_2SO_4$ in the product is 1:1. Therefore, 20.39 moles of H_2SO_4 will produce 20.39 moles of $(NH_4)_2SO_4$.

The maximum mass of ammonium sulphate that can be produced is calculated as follows:

$$m = n \times M = 20.41mol \times 132g/mol = 2694g$$

The maximum amount of ammonium sulphate that can be produced is 2.694 kg.

Exercise: Stoichiometry

1. Diborane, B_2H_6, was once considered for use as a rocket fuel. The combustion reaction for diborane is:

   $$B_2H_6(g) + 3O_2(l) \rightarrow 2HBO_2(g) + 2H_2O(l)$$

   If we react 2.37 grams of diborane, how many grams of water would we expect to produce?

2. Sodium azide is a commonly used compound in airbags. When triggered, it has the following reaction:

   $$2NaN_3(s) \rightarrow 2Na(s) + 3N_2(g)$$

   If 23.4 grams of sodium azide are reacted, how many moles of nitrogen gas would we expect to produce?

3. Photosynthesis is a chemical reaction that is vital to the existence of life on Earth. During photosynthesis, plants and bacteria convert carbon dioxide gas, liquid water, and light into glucose ($C_6H_{12}O_6$) and oxygen gas.
   (a) Write down the equation for the photosynthesis reaction.
   (b) Balance the equation.
   (c) If 3 moles of carbon dioxide are used up in the photosynthesis reaction, what mass of glucose will be produced?

13.9 Summary

- It is important to be able to quantify the changes that take place during a chemical reaction.
- The **mole (n)** is a SI unit that is used to describe an amount of substance that contains the same number of particles as there are atoms in 12 g of carbon.
- The number of particles in a mole is called the **Avogadro constant** and its value is 6.023 $\times$ 10^{23}. These particles could be atoms, molecules or other particle units, depending on the substance.
- The **molar mass (M)** is the mass of one mole of a substance and is measured in grams per mole or g.mol^{-1}. The numerical value of an element's molar mass is the same as its relative atomic mass. For a compound, the molar mass has the same numerical value as the molecular mass of that compound.
- The relationship between moles (n), mass in grams (m) and molar mass (M) is defined by the following equation:
$$n = \frac{m}{M}$$
- In a balanced chemical equation, the number in front of the chemical symbols describes the **mole ratio** of the reactants and products.
- The **empirical formula** of a compound is an expression of the relative number of each type of atom in the compound.
- The **molecular formula** of a compound describes the actual number of atoms of each element in a molecule of the compound.
- The formula of a substance can be used to calculate the **percentage by mass** that each element contributes to the compound.
- The **percentage composition** of a substance can be used to deduce its chemical formula.
- One mole of gas occupies a volume of 22.4 dm^3.
- The **concentration** of a solution can be calculated using the following equation,
$$C = \frac{n}{V}$$
where C is the concentration (in mol.dm^{-3}), n is the number of moles of solute dissolved in the solution and V is the volume of the solution (in dm^3).
- **Molarity** is a measure of the concentration of a solution, and its units are mol.dm^{-3}.
- **Stoichiometry**, the study of the relationships between reactants and products, can be used to determine the quantities of reactants and products that are involved in chemical reactions.

- A **limiting reagent** is the chemical that is used up first in a reaction, and which therefore determines how far the reaction will go before it has to stop.
- An **excess reagent** is a chemical that is in greater quantity than the limiting reagent in the reaction. Once the reaction is complete, there will still be some of this chemical that has not been used up.

Exercise: Summary Exercise

1. Write only the word/term for each of the following descriptions:
   (a) the mass of one mole of a substance
   (b) the number of particles in one mole of a substance
2. Multiple choice: Choose the one correct answer from those given.
   A 5 g of magnesium chloride is formed as the product of a chemical reaction. Select the **true** statement from the answers below:
      i. 0.08 moles of magnesium chloride are formed in the reaction
      ii. the number of atoms of Cl in the product is approximately 0.6023×10^{23}
      iii. the number of atoms of Mg is 0.05
      iv. the atomic ratio of Mg atoms to Cl atoms in the product is 1:1
   B 2 moles of oxygen gas react with hydrogen. What is the mass of oxygen in the reactants?
      i. 32 g
      ii. 0.125 g
      iii. 64 g
      iv. 0.063 g
   C In the compound potassium sulphate (K_2SO_4), oxygen makes up x% of the mass of the compound. x = ...
      i. 36.8
      ii. 9.2
      iii. 4
      iv. 18.3
   D The molarity of a 150 cm^3 solution, containing 5 g of NaCl is...
      i. 0.09 M
      ii. 5.7×10^{-4} M
      iii. 0.57 M
      iv. 0.03 M
3. 300 cm^3 of a 0.1 $mol.dm^{-3}$ solution of sulfuric acid is added to 200 cm^3 of a 0.5 $mol.dm^{-3}$ solution of sodium hydroxide.
   a Write down a balanced equation for the reaction which takes place when these two solutions are mixed.
   b Calculate the number of moles of sulfuric acid which were added to the sodium hydroxide solution.
   c Is the number of moles of sulfuric acid enough to fully neutralise the sodium hydroxide solution? Support your answer by showing all relevant calculations.
   (IEB Paper 2 2004)
4. Ozone (O_3) reacts with nitrogen monoxide gas (NO) to produce NO_2 gas. The NO gas forms largely as a result of emissions from the exhausts of motor vehicles and from certain jet planes. The NO_2 gas also causes the brown smog (smoke and fog), which is seen over most urban areas. This gas is also harmful to humans, as it causes breathing (respiratory) problems. The following equation indicates the reaction between ozone and nitrogen monoxide:

$$O_3(g) + NO(g) \rightarrow O_2(g) + NO_2(g)$$

In one such reaction 0.74 g of O_3 reacts with 0.67 g NO.

a Calculate the number of moles of O_3 and of NO present at the start of the reaction.

b Identify the limiting reagent in the reaction and justify your answer.

c Calculate the mass of NO_2 produced from the reaction.

(DoE Exemplar Paper 2, 2007)

5. A learner is asked to make 200 cm^3 of sodium hydroxide (NaOH) solution of concentration 0.5 $mol.dm^{-3}$.

   a Determine the mass of sodium hydroxide pellets he needs to use to do this.

   b Using an accurate balance the learner accurately measures the correct mass of the NaOH pellets. To the pellets he now adds exactly 200 cm^3 of pure water. Will his solution have the correct concentration? Explain your answer.
   300 cm^3 of a 0.1 $mol.dm^{-3}$ solution of sulfuric acid (H_2SO_4) is added to 200 cm^3 of a 0.5 $mol.dm^{-3}$ solution of NaOH at 25^0C.

   c Write down a balanced equation for the reaction which takes place when these two solutions are mixed.

   d Calculate the number of moles of H_2SO_4 which were added to the NaOH solution.

   e Is the number of moles of H_2SO_4 calculated in the previous question enough to fully neutralise the NaOH solution? Support your answer by showing all the relevant calculations.
   (IEB Paper 2, 2004)

Chapter 14

Energy Changes In Chemical Reactions - Grade 11

All chemical reactions involve energy changes. In some reactions, we are able to see these energy changes by either an increase or a decrease in the overall energy of the system.

14.1 What causes the energy changes in chemical reactions?

When a chemical reaction occurs, bonds in the reactants *break*, while new bonds *form* in the product. The following example may help to explain this.

Hydrogen reacts with oxygen to form water, according to the following equation:

$$2H_2 + O_2 \rightarrow 2H_2O$$

In this reaction, the bond between the two hydrogen atoms in the H_2 molecule will *break*, as will the bond between the oxygen atoms in the O_2 molecule. New bonds will *form* between the two hydrogen atoms and the single oxygen atom in the water molecule that is formed as the product.

For bonds to *break*, energy must be *absorbed*. When new bonds *form*, energy is *released*. The energy that is needed to break a bond is called the **bond energy** or **bond dissociation energy**. Bond energies are measured in units of kJ.mol^{-1}.

Definition: Bond energy
Bond energy is a measure of bond strength in a chemical bond. It is the amount of energy (in kJ.mol^{-1}) that is needed to break the chemical bond between two atoms.

14.2 Exothermic and endothermic reactions

In some reactions, the energy that must be *absorbed* to break the bonds in the reactants, is less than the total energy that is *released* when new bonds are formed. This means that in the overall reaction, energy is *released* as either heat or light. This type of reaction is called an **exothermic** reaction. Another way of describing an exothermic reaction is that it is one in which the energy of the product is less than the energy of the reactants, because energy has been released during the reaction. We can represent this using the following general formula:

$$\text{Reactants} \rightarrow \text{Product} + \text{Energy}$$

Definition: Exothermic reaction
An exothermic reaction is one that releases energy in the form of heat or light.

In other reactions,the energy that must be *absorbed* to break the bonds in the reactants, is more than the total energy that is *released* when new bonds are formed. This means that in the overall reaction, energy must be *absorbed* from the surroundings. This type of reaction is known as an **endothermic** reaction. Another way of describing an endothermic reaction is that it is one in which the energy of the product is greater than the energy of the reactants, because energy has been absorbed during the reaction. This can be represented by the following formula:

$$\text{Reactants} + \text{Energy} \rightarrow \text{Product}$$

Definition: Endothermic reaction
An endothermic reaction is one that absorbs energy in the form of heat.

The difference in energy (E) between the reactants and the products is known as the **heat of the reaction**. It is also sometimes referred to as the **enthalpy change** of the system.

Activity :: Demonstration : Endothermic and exothermic reactions 1
Apparatus and materials:
You will need citric acid, sodium bicarbonate, a glass beaker, the lid of an ice-cream container, thermometer, glass stirring rod and a pair of scissors. Note that citric acid is found in citrus fruits such as lemons. Sodium bicarbonate is actually bicarbonate of soda (baking soda), the baking ingredient that helps cakes to rise.
Method:

1. Cut a piece of plastic from the ice-cream container lid that will be big enough to cover the top of the beaker. Cut a small hole in the centre of this piece of plastic and place the thermometer through it.
2. Pour some citric acid ($H_3C_6H_5O_7$) into the glass beaker, cover the beaker with its 'lid' and record the temperature of the solution.
3. Stir in the sodium bicarbonate ($NaHCO_3$), then cover the beaker again.
4. Immediately record the temperature, and then take a temperature reading every two minutes after that. Record your results in a table like the one below.

Time (mins)	0	2	4	6
Temperature (^{0}C)				

The equation for the reaction that takes place is:

$$H_3C_6H_5O_7(aq) + 3NaHCO_3(s) \rightarrow 3CO_2(g) + 3H_2O(l) + NaC_6H_5O_7(aq)$$

Results:

- Plot your temperature results on a graph of temperature against time. What happens to the temperature during this reaction?
- Is this an exothermic or an endothermic reaction?
- Why was it important to keep the beaker covered with a lid?
- Do you think a glass beaker is the best thing to use for this experiment? Explain your answer.
- Suggest another container that could have been used and give reasons for your choice. It might help you to look back to chapter **??** for some ideas!

Activity :: Demonstration : Endothermic and exothermic reactions 2

Apparatus and materials:

Vinegar, steel wool, thermometer, glass beaker and plastic lid (from previous demonstration).

Method:

1. Put the thermometer through the plastic lid, cover the beaker and record the temperature in the empty beaker. You will need to leave the thermometer in the beaker for about 5 minutes in order to get an accurate reading.
2. Take the thermometer out of the jar.
3. Soak a piece of steel wool in vinegar for about a minute. The vinegar removes the protective coating from the steel wool so that the metal is exposed to oxygen.
4. After the steel wool has been in the vinegar, remove it and squeeze out any vinegar that is still on the wool. Wrap the steel wool around the thermometer and place it (still wrapped round the thermometer) back into the jar. The jar is automatically sealed when you do this because the thermometer is through the top of the lid.
5. Leave the steel wool in the beaker for about 5 minutes and then record the temperature. Record your observations.

Results:

You should notice that the temperature *increases* when the steel wool is wrapped around the thermometer.

Conclusion:

The reaction between oxygen and the exposed metal in the steel wool, is **exothermic**, which means that energy is released and the temperature increases.

14.3 The heat of reaction

The **heat of the reaction** is represented by the symbol ΔH, where:

$$\Delta H = E_{prod} - E_{react}$$

- In an *exothermic* reaction, ΔH is less than zero because the energy of the reactants is greater than the energy of the product. For example,

  $H_2 + Cl_2 \rightarrow 2HCl$ ΔH = -183 kJ

- In an *endothermic* reaction, ΔH is greater than zero because the energy of the reactants is less than the energy of the product. For example,

  $C + H_2O \rightarrow CO + H_2$ ΔH = +131 kJ

Some of the information relating to exothermic and endothermic reactions is summarised in table 14.1.

Table 14.1: A comparison of exothermic and endothermic reactions

Type of reaction	**Exothermic**	**Endothermic**
Energy absorbed or released	Released	Absorbed
Relative energy of reactants and products	Energy of reactants greater than energy of product	Energy of reactants less than energy of product
Sign of ΔH	Negative	Positive

Definition: Enthalpy
Enthalpy is the heat content of a chemical system, and is given the symbol 'H'.

Important: Writing equations using ΔH

There are two ways to write the heat of the reaction in an equation
For the exothermic reaction $C(s) + O_2(g) \rightarrow CO_2(g)$, we can write:

$C(s) + O_2(g) \rightarrow CO_2(g)$ ΔH = -393 kJ.mol^{-1} or
$C(s) + O_2(g) \rightarrow CO_2(g)$ + 393 kJ.mol^{-1}

For the endothermic reaction, $C(s) + H_2O(g) \rightarrow H_2(g) + CO(g)$, we can write:

$C(s) + H_2O(g) \rightarrow H_2(g) + CO(g)$ ΔH = +131 kJ.mol^{-1} or
$C(s) + H_2O(g)$ + 131 kJ.mol^{-1} $\rightarrow CO + H_2$

The **units** for ΔH are kJ.mol^{-1}. In other words, the ΔH value gives the amount of energy that is absorbed or released per mole of product that is formed. Units can also be written as kJ, which then gives the total amount of energy that is released or absorbed when the product forms.

Activity :: Investigation : Endothermic and exothermic reactions
Apparatus and materials:
Approximately 2 g each of calcium chloride ($CaCl_2$), sodium hydroxide (NaOH), potassium nitrate (KNO_3) and barium chloride ($BaCl_2$); concentrated sulfuric acid (H_2SO_4); 5 test tubes; thermometer.
Method:

1. Dissolve about 1 g of each of the following substances in 5-10 cm^3 of water in a test tube: $CaCl_2$, NaOH, KNO_3 and $BaCl_2$.
2. Observe whether the reaction is endothermic or exothermic, either by feeling whether the side of the test tube gets hot or cold, or using a thermometer.
3. Dilute 3 cm^3 of concentrated H_2SO_4 in 10 cm^3 of water in the fifth test tube and observe whether the temperature changes.
4. Wait a few minutes and then add NaOH to the H_2SO_4. Observe any energy changes.
5. Record which of the above reactions are endothermic and which are exothermic.

Results:

- When $BaCl_2$ and KNO_3 dissolve in water, they take in heat from the surroundings. The dissolution of these salts is **endothermic**.
- When $CaCl_2$ and NaOH dissolve in water, heat is released. The process is **exothermic**.

- The reaction of H_2SO_4 and NaOH is also **exothermic**.

14.4 Examples of endothermic and exothermic reactions

There are many examples of endothermic and exothermic reactions that occur around us all the time. The following are just a few examples.

1. **Endothermic reactions**

   - **Photosynthesis**
     Photosynthesis is the chemical reaction that takes place in plants, which uses energy from the sun to change carbon dioxide and water into food that the plant needs to survive, and which other organisms (such as humans and other animals) can eat so that they too can survive. The equation for this reaction is:

     $$6\mathrm{CO_2} + 12\mathrm{H_2O} + \mathrm{energy} \rightarrow \mathrm{C_6H_{12}O_6} + 6\mathrm{O_2} + 6\mathrm{H_2O}$$

     Photosynthesis is an endothermic reaction because it will not happen without an external source of energy, which in this case is sunlight.

   - **The thermal decomposition of limestone**
     In industry, the breakdown of limestone into quicklime and carbon dioxide is very important. Quicklime can be used to make steel from iron and also to neutralise soils that are too acid. However, the limestone must be heated in a kiln at a temperature of over 900^0C before the decomposition reaction will take place. The equation for the reaction is shown below:

     $$\mathrm{CaCO_3} \rightarrow \mathrm{CaO} + \mathrm{CO_2}$$

2. **Exothermic reactions**

   - **Combustion reactions** - The burning of fuel is an example of a combustion reaction, and we as humans rely heavily on this process for our energy requirements. The following equations describe the combustion of a hydrocarbon such as *methane* (CH_4):

     $$\mathrm{Fuel} + \mathrm{Oxygen} \rightarrow \mathrm{Heat} + \mathrm{Water} + \mathrm{CarbonDioxide}$$
     $$\mathrm{CH_4} + 2\mathrm{O_2} \rightarrow \mathrm{Heat} + \mathrm{H_2O} + \mathrm{CO_2}$$

     This is why we burn fuels for energy, because the chemical changes that take place during the reaction release huge amounts of energy, which we then use for things like power and electricity. You should also note that *carbon dioxide* is produced during this reaction. Later we will discuss some of the negative impacts of CO_2 on the environment. The chemical reaction that takes place when fuels burn therefore has both positive and negative consequences.

   - **Respiration**
     Respiration is the chemical reaction that happens in our bodies to produce energy for our cells. The equation below describes what happens during this reaction:

     $$\mathrm{C_6H_{12}O_6} + 6\mathrm{O_2} \rightarrow 6\mathrm{CO_2} + 6\mathrm{H_2O} + \mathrm{energy}$$

     In the reaction above, glucose (a type of carbohydrate in the food we eat) reacts with oxygen from the air that we breathe in, to form carbon dioxide (which we breathe out), water and energy. The energy that is produced allows the cell to carry out its functions efficiently. Can you see now why you are always told that you must eat food to get energy? It is not the food itself that provides you with energy, but the exothermic reaction that takes place when compounds within the food react with the oxygen you have breathed in!

Interesting Fact

Lightsticks or glowsticks are used by divers, campers, and for decoration and fun. A lightstick is a plastic tube with a glass vial inside it. To activate a lightstick, you bend the plastic stick, which breaks the glass vial. This allows the chemicals that are inside the glass to mix with the chemicals in the plastic tube. These two chemicals react and release energy. Another part of a lightstick is a fluorescent dye which changes this energy into light, causing the lightstick to glow!

Exercise: Endothermic and exothermic reactions

1. In each of the following reactions, say whether the reaction is endothermic or exothermic, and give a reason for your answer.
   (a) $H_2 + I_2 \rightarrow 2HI + 21kJ$
   (b) $CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O$ Δ H = -802 kJ
   (c) The following reaction takes place in a flask:
   $Ba(OH)_2.8H_2O + 2NH_4NO_3 \rightarrow Ba(NO_3)_2 + 2NH_3 + 10H_2O$
   Within a few minutes, the temperature of the flask drops by approximately 20C.
   (d) $Na + Cl_2 \rightarrow 2NaCl$ ΔH = -411 kJ
   (e) $C + O_2 \rightarrow CO_2$
2. For each of the following descriptions, say whether the process is endothermic or exothermic and give a reason for your answer.
   (a) evaporation
   (b) the combustion reaction in a car engine
   (c) bomb explosions
   (d) melting ice
   (e) digestion of food
   (f) condensation

14.5 Spontaneous and non-spontaneous reactions

Activity :: Demonstration : Spontaneous and non-spontaneous reactions
Apparatus and materials:
A length of magnesium ribbon, thick copper wire and a bunsen burner

Method:

1. Scrape the length of magnesium ribbon and copper wire clean.
2. Heat each piece of metal over the bunsen burner, in a non-luminous flame. Observe whether any chemical reaction takes place.
3. Remove the metals from the flame and observe whether the reaction stops. If the reaction stops, return the metal to the bunsen flame and continue to heat it.

Results:

- Did any reaction take place before the metals were heated?
- Did either of the reactions continue after they were removed from the flame?
- Write a balanced equation for each of the chemical reactions that takes place.

In the demonstration above, the reaction between magnesium and oxygen, and the reaction between copper and oxygen are both **non-spontaneous**. Before the metals were held over the bunsen burner, no reaction was observed. They need energy to *initiate* the reaction. After the reaction has started, it may then carry on spontaneously. This is what happened when the magnesium reacted with oxygen. Even after the magnesium was removed from the flame, the reaction continued. Other reactions will not carry on unless there is a constant addition of energy. This was the case when copper reacted with oxygen. As soon as the copper was removed from the flame, the reaction stopped.

Now try adding a solution of dilute sulfuric acid with a solution of sodium hydroxide. What do you observe? This is an example of a **spontaneous reaction** because the reaction takes place without any energy being added.

Definition: Spontaneous reaction
A spontaneous reaction is a physical or chemical change that occurs without the addition of energy.

14.6 Activation energy and the activated complex

From the demonstrations of spontaneous and non-spontaneous reactions, it should be clear that most reactions will not take place until the system has some minimum amount of energy added

to it. This energy is called the **activation energy**. **Activation energy** is the 'threshold energy' or the energy that must be overcome in order for a chemical reaction to occur.

Definition: Activation energy
Activation energy or 'threshold energy' is the energy that must be overcome in order for a chemical reaction to occur.

It is possible to draw an energy diagram to show the energy changes that take place during a particular reaction. Let's consider an example:

$$H_2(g) + F_2(g) \rightarrow 2HF(g)$$

Figure 14.1: The energy changes that take place during an exothermic reaction

The reaction between $H_2(g)$ and $F_2(g)$ (figure 14.1) needs energy in order to proceed, and this is the activation energy. Once the reaction has started, an in-between, temporary state is reached where the two reactants combine to give H_2F_2. This state is sometimes called a **transition state** and the energy that is needed to reach this state is equal to the activation energy for the reaction. The compound that is formed in this transition state is called the **activated complex**. The transition state lasts for only a very short time, after which either the original bonds reform, or the bonds are broken and a new product forms. In this example, the final product is HF and it has a lower energy than the reactants. The reaction is exothermic and ΔH is negative.

Definition: Activated complex
The activated complex is a transitional structure in a chemical reaction that results from the effective collisions between reactant molecules, and which remains while old bonds break and new bonds form.

In endothermic reactions, the final products have a higher energy than the reactants. An energy diagram is shown below (figure 14.2) for the endothermic reaction $XY + Z \rightarrow X + YZ$. In this example, the activated complex has the formula XYZ. Notice that the activation energy for the endothermic reaction is much greater than for the exothermic reaction.

Interesting Fact

The reaction between H and F was considered by NASA (National Aeronautics and Space Administration) as a fuel system for rocket boosters because of the energy that is released during this exothermic reaction.

Figure 14.2: The energy changes that take place during an endothermic reaction

Important: Enzymes and activation energy

An enzyme is a catalyst that helps to speed up the rate of a reaction by lowering the activation energy of a reaction. There are many enzymes in the human body, without which lots of important reactions would never take place. Cellular respiration is one example of a reaction that is catalysed by enzymes. You will learn more about catalysts in chapter **??**.

Exercise: Energy and reactions

1. Carbon reacts with water according to the following equation:

$$C + H_2O \Leftrightarrow CO + H_2 \ \Delta H > 0$$

   (a) Is this reaction endothermic or exothermic?
   (b) Give a reason for your answer.

2. Refer to the graph below and then answer the questions that follow:

(a) What is the energy of the reactants?
(b) What is the energy of the products?
(c) Calculate ΔH.
(d) What is the activation energy for this reaction?

14.7 Summary

- When a reaction occurs, some bonds break and new bonds form. These changes involve **energy**.
- When bonds break, energy is **absorbed** and when new bonds form, energy is **released**.
- The **bond energy** is the amount of energy that is needed to break the chemical bond between two atoms.
- If the energy that is needed to *break* the bonds is greater than the energy that is released when new bonds *form*, then the reaction is **endothermic**. The energy of the product is greater than the energy of the reactants.
- If the energy that is needed to *break* the bonds is less than the energy that is released when new bonds *form*, then the reaction is **exothermic**. The energy of the product is less than the energy of the reactants.
- An endothermic reaction is one that **absorbs** energy in the form of heat, while an exothermic reaction is one that **releases** energy in the form of heat and light.
- The difference in energy between the reactants and the product is called the **heat of reaction** and has the symbol ΔH.
- In an endothermic reaction, ΔH is a positive number, and in an exothermic reaction, ΔH will be negative.
- Photosynthesis, evaporation and the thermal decomposition of limestone, are all examples of endothermic reactions.
- Combustion reactions and respiration are both examples of exothermic reactions.
- A reaction which proceeds without additional energy being added, is called a **spontaneous reaction**.
- Reactions where energy must be *supplied* for the activation energy to be overcome, are called **non-spontaneous** reactions.
- In any reaction, some minimum energy must be overcome before the reaction will proceed. This is called the **activation energy** of the reaction.
- The **activated complex** is the transitional product that is formed during a chemical reaction while old bonds break and new bonds form.

Exercise: Summary Exercise

1. For each of the following, say whether the statement is **true** or **false**. If it is false, give a reason for your answer.
   (a) Energy is released in all chemical reactions.
   (b) The condensation of water vapour is an example of an endothermic reaction.

(c) In an exothermic reaction ΔH is less than zero.
(d) All non-spontaneous reactions are endothermic.

2. For each of the following, choose the one correct answer.
   (a) For the following reaction:
   $$A + B \Leftrightarrow AB \ \Delta H = \text{-129 kJ.mol}^{-1}$$
   i. The energy of the reactants is less than the energy of the product.
   ii. The energy of the product is less than the energy of the reactants.
   iii. The reaction is non-spontaneous.
   iv. The overall energy of the system increases during the reaction.
   (b) Consider the following chemical equilibrium:
   $$2NO_2 \Leftrightarrow N_2O_4$$
   Which one of the following graphs best represents the changes in potential energy that take place during the production of N_2O_4?

(i) (ii) (iii) (iv)

3. The cellular respiration reaction is catalysed by enzymes. The equation for the reaction is:
   $$C_6H_{12}O_6 + 6O_2 \rightarrow 6CO_2 + 6H_2O$$
   The change in potential energy during this reaction is shown below:

   (a) Will the value of ΔH be positive or negative? Give a reason for your answer.
   (b) Explain what is meant by 'activation energy'.
   (c) What role do enzymes play in this reaction?
   (d) Glucose is one of the reactants in cellular respiration. What important chemical reaction produces glucose?
   (e) Is the reaction in your answer above an endothermic or an exothermic one? Explain your answer.
   (f) Explain why proper nutrition and regular exercise are important in maintaining a healthy body.

Chapter 15

Types of Reactions - Grade 11

There are many different types of chemical reactions that can take place. In this chapter, we will be looking at a few of the more common reaction types: acid-base and acid-carbonate reactions, redox reactions and addition, elimination and substitution reactions.

15.1 Acid-base reactions

15.1.1 What are acids and bases?

In our daily lives, we encounter many examples of acids and bases. In the home, vinegar (acetic acid), lemon juice (citric acid) and tartaric acid (the main acid found in wine) are common, while hydrochloric acid, sulfuric acid and nitric acid are examples of acids that are more likely to be found in laboratories and industry. Hydrochloric acid is also found in the gastric juices in the stomach. Even fizzy drinks contain acid (carbonic acid), as do tea and wine (tannic acid)! Bases that you may have heard of include sodium hydroxide (caustic soda), ammonium hydroxide and ammonia. Some of these are found in household cleaning products. Acids and bases are also important commercial products in the fertiliser, plastics and petroleum refining industries. Some common acids and bases, and their chemical formulae, are shown in table 15.1.

Table 15.1: Some common acids and bases and their chemical formulae

Acid	**Formula**	**Base**	**Formula**
Hydrochoric acid	HCl	Sodium hydroxide	$NaOH$
Sulfuric acid	H_2SO_4	Potassium hydroxide	KOH
Nitric acid	HNO_3	Sodium carbonate	Na_2CO_3
Acetic (ethanoic) acid	CH_3COOH	Calcium hydroxide	$Ca(OH)_2$
Carbonic acid	H_2CO_3	Magnesium hydroxide	$Mg(OH)_2$
Sulfurous acid	H_2SO_3	Ammonia	NH_3
Phosphoric acid	H_3PO_4	Sodium bicarbonate	$NaHCO_3$

Most acids share certain characteristics, and most bases also share similar characteristics. It is important to be able to have a definition for acids and bases so that they can be correctly identified in reactions.

15.1.2 Defining acids and bases

A number of definitions for acids and bases have developed over the years. One of the earliest was the **Arrhenius** definition. Arrhenius (1887) noticed that water dissociates (splits up) into hydronium (H_3O^+) and hydroxide (OH^-) ions according to the following equation:

$$H_2O \Leftrightarrow H_3O^+ + OH^-$$

Arrhenius described an acid as a compound that increases the concentration of H_3O^+ ions in solution, and a base as a compound that increases the concentration of OH^- ions in a solution. Look at the following examples showing the dissociation of hydrochloric acid and sodium hydroxide (a base) respectively:

For more information on *dissociation*, refer to chapter 20.

1. $\mathrm{HCl} + \mathrm{H_2O} \rightarrow \mathrm{H_3O^+} + \mathrm{Cl^-}$

   Hydrochloric acid in water increases the concentration of H_3O^+ ions and is therefore an *acid*.

2. $\mathrm{NaOH} + \mathrm{H_2O} \rightarrow \mathrm{Na^+} + \mathrm{OH^-}$

   Sodium hydroxide in water increases the concentration of OH^- ions and is therefore a *base*.

However, this definition could only be used for acids and bases *in water*. It was important to come up with a much broader definition for acids and bases.

It was Lowry and Bronsted (1923) who took the work of Arrhenius further to develop a broader definition for acids and bases. The **Bronsted-Lowry model** defines acids and bases in terms of their ability to donate or accept protons.

Definition: Acids and bases
According to the Bronsted-Lowry theory of acids and bases, an **acid** is a substance that gives away protons (H^+), and is therefore called a **proton donor**. A **base** is a substance that takes up protons, and is therefore called a **proton acceptor**.

Below are some examples:

1. $\mathrm{HCl(g)} + \mathrm{NH_3(g)} \rightarrow \mathrm{NH^{4+}} + \mathrm{Cl^-}$

   In order to decide which substance is a proton donor and which is a proton acceptor, we need to look at what happens to each reactant. The reaction can be broken down as follows:

   $\mathrm{HCl} \rightarrow \mathrm{Cl^-} + \mathrm{H^+}$ and
   $\mathrm{NH_3} + \mathrm{H^+} \rightarrow \mathrm{NH_4^+}$

   From these reactions, it is clear that HCl is a *proton donor* and is therefore an **acid**, and that NH_3 is a *proton acceptor* and is therefore a **base**.

2. $\mathrm{CH_3COOH} + \mathrm{H_2O} \rightarrow \mathrm{H_3O^+} + \mathrm{CH_3COO^-}$

   The reaction can be broken down as follows:

   $\mathrm{CH_3COOH} \rightarrow \mathrm{CH_3COO^-} + \mathrm{H^+}$ and
   $\mathrm{H_2O} + \mathrm{H^+} \rightarrow \mathrm{H_3O^+}$

   In this reaction, CH_3COOH (acetic acid) is a proton donor and is therefore the **acid**. In this case, water acts as a base because it accepts a proton to form H_3O^+.

3. $\mathrm{NH_3} + \mathrm{H_2O} \rightarrow \mathrm{NH_4^+} + \mathrm{OH^-}$

   The reaction can be broken down as follows:

$H_2O \rightarrow OH^- + H^+$ and

$NH_3 + H^+ \rightarrow NH_4^+$

In this reaction, water donates a proton and is therefore an **acid** in this reaction. Ammonia accepts the proton and is therefore the **base**. Notice that in the previous equation, water acted as a base and that in this equation it acts as an acid. Water can act as both an acid and a base depending on the reaction. This is also true of other substances. These substances are called **ampholytes** and are said to be **amphoteric**.

Definition: Amphoteric
An amphoteric substance is one that can react as either an acid or base. Examples of amphoteric substances include water, zinc oxide and beryllium hydroxide.

15.1.3 Conjugate acid-base pairs

Look at the reaction between hydrochloric acid and ammonia to form ammonium and chloride ions:

$$HCl + NH_3 \Leftrightarrow NH_4^+ + Cl^-$$

Looking firstly at the *forward reaction* (i.e. the reaction that proceeds from *left* to *right*), the changes that take place can be shown as follows:

$HCl \rightarrow Cl^- + H^+$ and

$NH_3 + H^+ \rightarrow NH_4^+$

Looking at the *reverse reaction* (i.e. the reaction that proceeds from *right* to *left*), the changes that take place are as follows:

$NH_4^+ \rightarrow NH_3 + H^+$ and

$Cl^- + H^+ \rightarrow HCl$

In the **forward reaction**, HCl is a proton donor (acid) and NH_3 is a proton acceptor (base). In the **reverse reaction**, the chloride ion is the proton acceptor (base) and NH_4^+ is the proton donor (acid). A **conjugate acid-base pair** is two compounds in a reaction that change into each other through the loss or gain of a proton. The conjugate acid-base pairs for the above reaction are shown below.

The reaction between ammonia and water can also be used as an example:

conjugate pair

$$\underset{\text{acid}_1}{\mathbf{H_2O}} + \underset{\text{base}_2}{\mathbf{NH_3}} \rightleftharpoons \underset{\text{acid}_2}{\mathbf{NH_4^+}} + \underset{\text{base}_1}{\mathbf{OH^-}}$$

conjugate pair

Definition: Conjugate acid-base pair
The term refers to two compounds that transform into each other by the gain or loss of a proton.

Exercise: Acids and bases

1. In the following reactions, identify (1) the acid and the base in the reactants and (2) the salt in the product.
   (a) $H_2SO_4 + Ca(OH)_2 \rightarrow CaSO_4 + 2H_2O$
   (b) $CuO + H_2SO_4 \rightarrow CuSO_4 + H_2O$
   (c) $H_2O + C_6H_5OH \rightarrow H_3O^+ + C_6H_5O^-$
   (d) $HBr + C_5H_5N \rightarrow (C_5H_5NH^+)Br^-$
2. In each of the following reactions, label the conjugate acid-base pairs.
   (a) $H_2SO_4 + H_2O \Leftrightarrow H_3O^+ + HSO_4^-$
   (b) $NH_4^+ + F^- \Leftrightarrow HF + NH_3$
   (c) $H_2O + CH_3COO^- \Leftrightarrow CH_3COOH + OH^-$
   (d) $H_2SO_4 + Cl^- \Leftrightarrow HCl + HSO_4^-$

15.1.4 Acid-base reactions

When an acid and a base react, they neutralise each other to form a **salt**. If the base contains hydroxide (OH^-) ions, then *water* will also be formed. The word *salt* is a general term which applies to the products of all acid-base reactions. A salt is a product that is made up of the cation from a base and the anion from an acid. When an acid reacts with a base, they **neutralise** each other. In other words, the acid becomes less acidic and the base becomes less basic. Look at the following examples:

1. Hydrochloric acid reacts with sodium hydroxide to form sodium chloride (the salt) and water. Sodium chloride is made up of Na^+ cations from the base (NaOH) and Cl^- anions from the acid (HCl).

   $HCl + NaOH \rightarrow H_2O + NaCl$

2. Hydrogen bromide reacts with potassium hydroxide to form potassium bromide (the salt) and water. Potassium bromide is made up of K^+ cations from the base (KOH) and Br^- anions from the acid (HBr).

   $HBr + KOH \rightarrow H_2O + KBr$

3. Hydrochloric acid reacts with sodium hydrocarbonate to form sodium chloride (the salt) and hydrogen carbonate. Sodium chloride is made up of Na^+ cations from the base ($NaHCO_3$) and Cl^- anions from the acid (HCl).

$$\mathrm{HCl} + \mathrm{NaHCO_3} \rightarrow \mathrm{H_2CO_3} + \mathrm{NaCl}$$

You should notice that in the first two examples, the base contained OH^- ions, and therefore the products were a *salt* and *water*. NaCl (table salt) and KBr are both salts. In the third example, $NaHCO_3$ also acts as a base, despite not having OH^- ions. A salt is still formed as one of the products, but no water is produced.

It is important to realise how important these neutralisation reactions are. Below are some examples:

- **Domestic uses**

  Calcium oxide (CaO) is put on soil that is too acid. Powdered limestone ($CaCO_3$) can also be used but its action is much slower and less effective. These substances can also be used on a larger scale in farming and also in rivers.

- **Biological uses**

  Acids in the stomach (e.g. hydrochloric acid) play an important role in helping to digest food. However, when a person has a stomach ulcer, or when there is too much acid in the stomach, these acids can cause a lot of pain. **Antacids** are taken to neutralise the acids so that they don't burn as much. Antacids are bases which neutralise the acid. Examples of antacids are aluminium hydroxide, magnesium hydroxide ('milk of magnesia') and sodium bicarbonate ('bicarbonate of soda'). Antacids can also be used to relieve heartburn.

- **Industrial uses**

  Alkaline calcium hydroxide (limewater) can be used to absorb harmful SO_2 gas that is released from power stations and from the burning of fossil fuels.

Interesting Fact

Bee stings are acidic and have a pH between 5 and 5.5. They can be soothed by using substances such as calomine lotion, which is a mild alkali based on zinc oxide. Bicarbonate of soda can also be used. Both alkalis help to neutralise the acidic bee sting and relieve some of the itchiness!

Important: Acid-base titrations

The neutralisation reaction between an acid and a base can be very useful. If an acidic solution of known concentration (a standard solution) is added to an alkaline solution until the solution is exactly neutralised (i.e. it has neither acidic nor basic properties), it is possible to calculate the exact concentration of the unknown solution. It is possible to do this because, at the exact point where the solution is neutralised, chemically equivalent amounts of acid and base have reacted with each other. This type of calculation is called **volumetric analysis**. The process where an acid solution and a basic solution are added to each other for this purpose, is called a **titration**, and the point of neutralisation is called the **end point** of the reaction. So how exactly can a titration be carried out to determine an unknown concentration? Look at the following steps to help you to understand the process.

Step 1:
A measured volume of the solution with unknown concentration is put into a flask.

Step 2:
A suitable indicator is added to this solution (bromothymol blue and phenolpthalein are common indicators).

Step 3:
A volume of the standard solution is put into a burette and is slowly added to the solution in the flask, drop by drop.

Step 4:
At some point, adding one more drop will change the colour of the unknown solution. For example, if the solution is basic and bromothymol blue is being used as the indicator in the titration, the bromothymol blue would originally have coloured the solution blue. At the end point of the reaction, adding one more drop of acid will change the colour of the basic solution from blue to yellow. Yellow shows that the solution is now acidic.

Step 5:
Record the volume of standard solution that has been added up to this point.

Step 6:
Use the information you have gathered to calculate the exact concentration of the unknown solution. A worked example is shown below.

Important: Titration calculations

When you are busy with these calculations, you will need to remember the following:

1dm^3 = 1 litre = 1000ml = 1000cm^3, therefore dividing cm^3 by 1000 will give you an answer in dm^3.

Some other terms and equations which will be useful to remember are shown below:

- Molarity is a term used to describe the concentration of a solution, and is measured in mol.dm^{-3}. The symbol for molarity is M. Refer to chapter 13 for more information on molarity.
- Moles = molarity (mol.dm^{-3}) x volume (dm^3)
- Molarity (mol.dm^{-3}) = mol / volume

Worked Example 74: Titration calculation

Question: Given the equation:

$$\mathrm{NaOH} + \mathrm{HCl} \rightarrow \mathrm{NaCl} + \mathrm{H_2O}$$

25cm^3 of a sodium hydroxide solution was pipetted into a conical flask and titrated with 0.2M hydrochloric acid. Using a suitable indicator, it was found that 15cm^3 of acid was needed to neutralise the alkali. Calculate the molarity of the sodium hydroxide.

Answer
Step 1 : Write down all the information you know about the reaction, and make sure that the equation is balanced.
NaOH: V = 25 cm^3
HCl: V = 15 cm^3; C = 0.2 M
The equation is already balanced.

Step 2 : Calculate the number of moles of HCl that react according to this equation.

$$M = \frac{n}{V}$$

Therefore, n(HCl) = M × V (make sure that all the units are correct!)

M = 0.2mol.dm^{-3}
V = 15cm^3 = 0.015dm^3

Therefore

$$n(HCl) = 0.2 \times 0.015 = 0.003$$

There are 0.003 moles of HCl that react

Step 3 : Calculate the number of moles of sodium hydroxide in the reaction
Look at the equation for the reaction. For every mole of HCl there is one mole of NaOH that is involved in the reaction. Therefore, if 0.003 moles of HCl react, we can conclude that the same quantity of NaOH is needed for the reaction. The number of moles of NaOH in the reaction is 0.003.

Step 4 : Calculate the molarity of the sodium hydroxide
First convert the volume into dm^3. V = 0.025 dm^3. Then continue with the calculation.

$$M = \frac{n}{V} = \frac{0.003}{0.025} = 0.12$$

The molarity of the NaOH solution is 0.12 mol.dm^3 or 0.12 M

Worked Example 75: Titration calculation

Question: 4.9 g of sulfuric acid is dissolved in water and the final solution has a volume of 220 cm^3. Using titration, it was found that 20 cm^3 of this solution was

able to completely neutralise 10 cm^3 of a sodium hydroxide solution. Calculate the concentration of the sodium hydroxide in $mol.dm^{-3}$.

Answer

Step 1 : Write a balanced equation for the titration reaction.

$H_2SO_4 + 2NaOH \rightarrow Na_2SO_4 + 2H_2O$

Step 2 : Calculate the molarity of the sulfuric acid solution.

M = n/V

V = 220 cm^3 = 0.22 dm^3

$$n = \frac{m}{M} = \frac{4.9g}{98g.mol^{-1}} = 0.05mols$$

Therefore,

$$M = \frac{0.05}{0.22} = 0.23mol.dm^{-3}$$

Step 3 : Calculate the moles of sulfuric acid that were used in the neutralisation reaction.

Remember that only 20 cm^3 of the sulfuric acid solution is used.

M = n/V, therefore n = M × V

$$n = 0.23 \times 0.02 = 0.0046mol$$

Step 4 : Calculate the number of moles of sodium hydroxide that were neutralised.

According to the balanced chemical equation, the mole ratio of H_2SO_4 to NaOH is 1:2. Therefore, the number of moles of NaOH that are neutralised is 0.0046 × 2 = 0.0092 mols.

Step 5 : Calculate the concentration of the sodium hydroxide solution.

$$M = \frac{n}{V} = \frac{0.0092}{0.01} = 0.92M$$

15.1.5 Acid-carbonate reactions

Activity :: Demonstration : The reaction of acids with carbonates

Apparatus and materials:

Small amounts of sodium carbonate and calcium carbonate (both in powder form); hydrochloric acid and sulfuric acid; retort stand; two test tubes; two rubber stoppers for the test tubes; a delivery tube; lime water. The demonstration should be set up as shown below.

Method:

1. Pour limewater into one of the test tubes and seal with a rubber stopper.
2. Pour a small amount of hydrochloric acid into the remaining test tube.
3. Add a small amount of sodium carbonate to the acid and seal the test tube with the rubber stopper.
4. Connect the two test tubes with a delivery tube.
5. Observe what happens to the colour of the limewater.
6. Repeat the above steps, this time using sulfuric acid and calcium carbonate.

Observations:
The clear lime water turns milky meaning that carbon dioxide has been produced.

When an acid reacts with a carbonate a salt, carbon dioxide and water are formed. Look at the following examples:

- Nitric acid reacts with sodium carbonate to form sodium nitrate, carbon dioxide and water.

$$2HNO_3 + Na_2CO_3 \rightarrow 2NaNO_3 + CO_2 + H_2O$$

- Sulfuric acid reacts with calcium carbonate to form calcium sulphate, carbon dioxide and water.

$$H_2SO_4 + CaCO_3 \rightarrow CaSO_4 + CO_2 + H_2O$$

- Hydrochloric acid reacts with calcium carbonate to form calcium chloride, carbon dioxide and water.

$$2HCl + CaCO_3 \rightarrow CaCl_2 + CO_2 + H_2O$$

?

Exercise: Acids and bases

1. The compound $NaHCO_3$ is commonly known as baking soda. A recipe requires 1.6 g of baking soda, mixed with other ingredients, to bake a cake.

(a) Calculate the number of moles of $NaHCO_3$ used to bake the cake.

(b) How many atoms of oxygen are there in the 1.6 g of baking soda?
During the baking process, baking soda reacts with an acid to produce carbon dioxide and water, as shown by the reaction equation below:

$$HCO_3^-(aq) + H^+(aq) \rightarrow CO_2(g) + H_2O(l)$$

(c) Identify the reactant which acts as the Bronsted-Lowry base in this reaction. Give a reason for your answer.

(d) Use the above equation to explain why the cake rises during this baking process.

(DoE Grade 11 Paper 2, 2007)

2. Label the acid-base conjugate pairs in the following equation:

$$HCO_3^- + H_2O \Leftrightarrow CO_3^{2-} + H_3O^+$$

3. A certain antacid tablet contains 22.0 g of baking soda ($NaHCO_3$). It is used to neutralise the excess hydrochloric acid in the stomach. The balanced equation for the reaction is:

$$NaHCO_3 + HCl \rightarrow NaCl + H_2O + CO_2$$

The hydrochloric acid in the stomach has a concentration of 1.0 mol.dm^{-3}. Calculate the volume of the hydrochloric acid that can be neutralised by the antacid tablet.

(DoE Grade 11 Paper 2, 2007)

4. A learner is asked to prepare a *standard solution* of the weak acid, oxalic acid $(COOH)_2 2H_2O$ for use in a titration. The volume of the solution must be 500 cm^3 and the concentration 0.2 mol.dm^{-3}.

(a) Calculate the mass of oxalic acid which the learner has to dissolve to make up the required standard solution. The leaner titrates this 0.2 mol.dm^{-3} oxalic acid solution against a solution of sodium hydroxide. He finds that 40 cm^3 of the oxalic acid solution exactlt neutralises 35 cm^3 of the sodium hydroxide solution.

(b) Calculate the concentration of the sodium hydroxide solution.

5. A learner finds some sulfuric acid solution in a bottle labelled 'dilute sulfuric acid'. He wants to determine the concentration of the sulphuric acid solution. To do this, he decides to titrate the sulphuric acid against a standard potassium hydroxide (KOH) solution.

(a) What is a standard solution?

(b) Calculate the mass of KOH which he must use to make 300 cm^3 of a 0.2 mol.dm^{-3} KOH solution.

(c) Calculate the pH of the 0.2 mol.dm^{-3} KOH solution (assume standard temperature).

(d) Write a balanced chemical equation for the reaction between H_2SO_4 and KOH.

(e) During the titration he finds that 15 cm^3 of the KOH solution neutralises 20 cm^3 of the H_2SO_4 solution. Calculate the concentration of the H_2SO_4 solution.

(IEB Paper 2, 2003)

15.2 Redox reactions

A second type of reaction is the **redox** reaction, in which both **oxidation** and **reduction** take place.

15.2.1 Oxidation and reduction

If you look back to chapter 4, you will remember that we discussed how, during a chemical reaction, an exchange of electrons takes place between the elements that are involved. Using *oxidation numbers* is one way of tracking what is happening to these electrons in a reaction. Refer back to section 4.11 if you can't remember the rules that are used to give an oxidation number to an element. Below are some examples to refresh your memory before we carry on with this section!

Examples:

1. **CO_2**

   Each oxygen atom has an oxidation number of -2. This means that the charge on *two* oxygen atoms is -4. We know that the molecule of CO_2 is neutral, therefore the carbon atom must have an oxidation number of +4.

2. **$KMnO_4$**

   Overall, this molecule has a neutral charge, meaning that the sum of the oxidation numbers of the elements in the molecule must equal zero. Potassium (K) has an oxidation number of +1, while oxygen (O) has an oxidation number of -2. If we exclude the atom of manganese (Mn), then the sum of the oxidation numbers equals +1+(-2x4)= -7. The atom of manganese must therefore have an oxidation number of +7 in order to make the molecule neutral.

By looking at how the oxidation number of an element changes during a reaction, we can easily see whether that element is being **oxidised** or **reduced**.

Definition: Oxidation and reduction
Oxidation is the *loses* of an electron by a molecule, atom or ion. Reduction is the *gain* of an electron by a molecule, atom or ion.

Example:

$$\mathrm{Mg} + \mathrm{Cl_2} \rightarrow \mathrm{MgCl_2}$$

As a *reactant*, magnesium has an oxidation number of zero, but as part of the *product* magnesium chloride, the element has an oxidation number of +2. Magnesium has *lost* two electrons and has therefore been *oxidised*. This can be written as a **half-reaction**. The half-reaction for this change is:

$$\mathrm{Mg} \rightarrow \mathrm{Mg^{2+}} + 2\mathrm{e^-}$$

As a *reactant*, chlorine has an oxidation number of zero, but as part of the *product* magnesium chloride, the element has an oxidation number of -1. Each chlorine atom has *gained* an electron and the element has therefore been *reduced*. The half-reaction for this change is:

$$\mathrm{Cl_2} + 2\mathrm{e^-} \rightarrow 2\mathrm{Cl^-}$$

Definition: Half-reaction
A half reaction is either the oxidation or reduction reaction part of a redox reaction. A half reaction is obtained by considering the change in oxidation states of the individual substances that are involved in the redox reaction.

Important: Oxidation and reduction made easy!

An easy way to think about oxidation and reduction is to remember:

'OILRIG' - **O**xidation **I**s **L**oss of electrons, **R**eduction **I**s **G**ain of electrons.

An element that is **oxidised** is called a **reducing agent**, while an element that is **reduced** is called an **oxidising agent**.

15.2.2 Redox reactions

Definition: Redox reaction
A redox reaction is one involving oxidation and reduction, where there is always a change in the oxidation numbers of the elements involved.

Activity :: Demonstration : Redox reactions
Materials:
A few granules of zinc; 15 ml copper (II) sulphate solution (blue colour), glass beaker.

Method:
Add the zinc granules to the copper sulphate solution and observe what happens. What happens to the zinc granules? What happens to the colour of the solution?
Results:

- Zinc becomes covered in a layer that looks like copper.
- The blue copper sulphate solution becomes clearer.

Cu^{2+} ions from the $CuSO_4$ solution are **reduced** to form copper metal. This is what you saw on the zinc crystals. The reduction of the copper ions (in other words, their removal from the copper sulphate solution), also explains the change in colour of the solution. The equation for this reaction is:

$$\mathrm{Cu^{2+} + 2e^- \rightarrow Cu}$$

Zinc is **oxidised** to form Zn^{2+} ions which are clear in the solution. The equation for this reaction is:

$$\mathrm{Zn \rightarrow Zn^{2+} + 2e^-}$$

The overall reaction is:

$$\mathrm{Cu^{2+}(aq) + Zn(s) \rightarrow Cu(s) + Zn^{2+}(aq)}$$

Conclusion:
A redox reaction has taken place. Cu^{2+} ions are reduced and the zinc is oxidised.

Below are some further examples of redox reactions:

- $H_2 + F_2 \rightarrow 2HF$ can be re-written as two half-reactions:

$$H_2 \rightarrow 2H^+ + 2e^- \text{ (oxidation) and}$$
$$F_2 + 2e^- \rightarrow 2F^- \text{ (reduction)}$$

- $Cl_2 + 2KI \rightarrow 2KCl + I_2$ or $Cl_2 + 2I^- \rightarrow 2Cl^- + I_2$, can be written as two half-reactions:

$$Cl_2 + 2e^- \rightarrow 2Cl^- \text{ (reduction) and}$$
$$2I^- \rightarrow I_2 + 2e^- \text{ (oxidation)}$$

In Grade 12, you will go on to look at electrochemical reactions, and the role that electron transfer plays in this type of reaction.

Exercise: Redox Reactions

1. Look at the following reaction:

$$2H_2O_2(l) \rightarrow 2H_2O(l) + O_2(g)$$

   (a) What is the oxidation number of the oxygen atom in each of the following compounds?
      i. H_2O_2
      ii. H_2O
      iii. O_2

   (b) Does the hydrogen peroxide (H_2O_2) act as an oxidising agent or a reducing agent or both, in the above reaction? Give a reason for your answer.

2. Consider the following chemical equations:
   1. $Fe(s) \rightarrow Fe^{2+}(aq) + 2e^-$
   2. $4H^+(aq) + O_2(g) + 4e^- \rightarrow 2H_2O(l)$

   Which one of the following statements is correct?

   (a) Fe is oxidised and H^+ is reduced
   (b) Fe is reduced and O_2 is oxidised
   (c) Fe is oxidised and O_2 is reduced
   (d) Fe is reduced and H^+ is oxidised

   (DoE Grade 11 Paper 2, 2007)

3. Which one of the following reactions is a redox reaction?

   (a) $HCl + NaOH \rightarrow NaCl + H_2O$
   (b) $AgNO_3 + NaI \rightarrow AgI + NaNO_3$
   (c) $2FeCl_3 + 2H_2O + SO_2 \rightarrow H_2SO_4 + 2HCl + 2FeCl_2$
   (d) $BaCl_2 + MgSO_4 \rightarrow MgCl_2 + BaSO_4$

15.3 Addition, substitution and elimination reactions

15.3.1 Addition reactions

An addition reaction occurs when two or more reactants combine to form a final product. This product will contain *all* the atoms that were present in the reactants. The following is a general equation for this type of reaction:

$$A + B \rightarrow C$$

Notice that C is the final product with no A or B remaining as a residue.

The following are some examples.

1. The reaction between **ethene** and **bromine** to form 1,2-dibromoethane (figure 15.1).

$$\mathrm{C_2H_4 + Br_2 \rightarrow C_2H_4Br_2}$$

Figure 15.1: The reaction between ethene and bromine is an example of an addition reaction

2. **Polymerisation reactions**

   In industry, making polymers is very important. A polymer is made up of lots of smaller units called *monomers*. When these monomers are added together, they form a polymer. Examples of polymers are polyvinylchloride (PVC) and polystyrene. PVC is often used to make piping, while polystyrene is an important packaging and insulating material. Polystyrene is made up of lots of styrene monomers which are joined through addition reactions (figure 15.2). 'Polymerisation' refers to the addition reactions that eventually help to form the polystyrene polymer.

Figure 15.2: The polymerisation of a styrene monomer to form a polystyrene polymer

3. The **hydrogenation** of vegetable oils to form margarine is another example of an addition reaction. Hydrogenation involves adding hydrogen (H_2) to an alkene. An alkene is an organic compound composed of carbon and hydrogen. It contains a double bond between two of the carbon atoms. If this bond is broken, it means that more hydrogen atoms can attach themselves to the carbon atoms. During hydrogenation, this double bond is broken, and more hydrogen atoms are added to the molecule. The reaction that takes place is shown below. Note that the 'R' represents any side-chain. A side-chain is simply any combination of atoms that are attached to the central part of the molecule.

$$\mathrm{RCHCH_2 + H_2 \rightarrow RCH_2CH_3}$$

4. The production of the alcohol ethanol from ethene. Ethanol (CH_3CH_2OH) can be made from alkenes such as ethene (C_2H_4), through a hydration reaction like the one below. A hydration reaction is one where water is added to the reactants.

$$C_2H_4 + H_2O \rightarrow CH_3CH_2OH$$

A catalyst is needed for this reaction to take place. The catalyst that is most commonly used is phosphoric acid.

15.3.2 Elimination reactions

An elimination reaction occurs when a reactant is broken up into two products. The general form of the equation is as follows:

$$A \rightarrow B + C$$

The examples below will help to explain this:

1. The **dehydration of an alcohol** is one example. Two hydrogen atoms and one oxygen atom are eliminated and a molecule of water is formed as a second product in the reaction, along with an alkene.

$$\mathrm{CH_3CH_2OH \rightarrow CH_2CH_2 + H_2O}$$

2. The elimination of **potassium bromide from a bromoalkane**.

$$\mathrm{CH_3CH_2Br + KOH \rightarrow CH_2CH_2 + KBr + H_2O}$$

3. Ethane cracking is an important industrial process used by SASOL and other petrochemical industries. Hydrogen is eliminated from ethane (C_2H_6) to produce an alkene called ethene (C_2H_4). Ethene is then used to produce other products such as polyethylene. You will learn more about these compounds in chapter 23. The equation for the cracking of ethane looks like this:

$$\mathrm{C_2H_6} \rightarrow \mathrm{C_2H_4} + \mathrm{H_2}$$

15.3.3 Substitution reactions

A substitution reaction occurs when an exchange of reactants takes place. The initial reactants are transformed or 'swopped around' to give a final product. A simple example of a reaction like this is shown below:

$$AB + CD \rightarrow AC + BD$$

Some simple examples of substitution reactions are shown below:

$$CH_4 + Cl_2 \rightarrow CH_3Cl + HCl$$

In this example, a chorine atom and a hydrogen atom are exchanged to create a new product.

$$Cu(H_2O)_4^{2+} + 4Cl^- \Leftrightarrow Cu(Cl)_4^{2-} + 4H_2O$$

Exercise: Addition, substitution and elimination reactions

1. Refer to the diagram below and then answer the questions that follow:

   (a) Is this reaction an example of substitution, elimination or addition?
   (b) Give a reason for your answer above.
   (c) What type of compound is the reactant marked (i)?

2. The following diagram shows the reactants in an addition reaction.

   (a) Draw the final product in this reaction.
   (b) What is the chemical formula of the product?

3. The following reaction takes place:

Is this reaction a substitution, addition or dehydration reaction? Give a reason for your answer.

4. Consider the following reaction:

$$\mathrm{Ca(OH)_2(s) + 2NH_4Cl(s) \rightarrow CaCl_2(s) + 2NH_3(g) + 2H_2O(g)}$$

Which one of the following best describes the type of reaction which takes place?

(a) Redox reaction
(b) Acid-base reaction
(c) Dehydration reaction

15.4 Summary

- There are many different **types of reactions** that can take place. These include acid-base, acid-carbonate, redox, addition, substitution and elimination reactions.
- The **Arrhenius** definition of acids and bases defines an acid as a substance that increases the concentration of hydrogen ions (H^+ or H_3O^+) in a solution. A base is a substance that increases the concentration of hydroxide ions (OH^-) in a solution. However this definition only applies to substances that are in water.
- The **Bronsted-Lowry** definition is a much broader one. An **acid** is a substance that **donates protons** and a **base** is a substance that **accepts protons**.
- In different reactions, certain substances can act as both an acid and a base. These substances are called **ampholytes** and are said to be **amphoteric**. Water is an example of an amphoteric substance.
- A **conjugate acid-base pair** refers to two compounds in a reaction that change into other through the loss or gain of a proton.
- When an acid and a base react, they form a **salt** and water. The salt is made up of a cation from the base and an anion from the acid. An example of a salt is sodium chloride (NaCl), which is the product of the reaction between sodium hydroxide (NaOH) and hydrochloric acid (HCl).
- The reaction between an acid and a base is a **neutralisation** reaction.
- **Titrations** are reactions between an acid and a base that are used to calculate the concentration of one of the reacting substances. The concentration of the other reacting substance must be known.
- In an **acid-carbonate reaction**, an acid and a carbonate react to form a salt, carbon dioxide and water.
- A **redox reaction** is one where there is always a change in the oxidation numbers of the elements that are involved in the reaction.
- **Oxidation** is the loss of electrons and **reduction** is the gain of electrons.

- When two or more reactants combine to form a product that contains all the atoms that were in the reactants, then this is an **addition reaction**. Examples of addition reactions include the reaction between ethene and bromine, polymerisation reactions and hydrogenation reactions.
- A reaction where the reactant is broken down into one or more product, is called an **elimination reaction**. Alcohol dehydration and ethane cracking are examples of elimination reactions.
- A **substitution reaction** is one where the reactants are transformed or swopped around to form the final product.

Exercise: Summary Exercise

1. Give one word/term for each of the following descriptions:
   (a) A chemical reaction during which electrons are transferred
   (b) The addition of hydrogen across a double bond
   (c) The removal of hydrogen and halogen atoms from an alkane to form an elkene
2. For each of the following, say whether the statement is true or false. If the statement is false, re-write the statement correctly.
   (a) The conjugate base of NH_4^+ is NH_3.
   (b) The reactions $C + O_2 \rightarrow CO_2$ and $2KClO_3 \rightarrow 2KCl + 3O_2$ are examples of redox reactions.
3. For each of the following questions, choose the one correct statement from the list provided.
   A The following chemical equation represents the formation of the hydronium ion:
   $$H^+(aq) + H_2O(l) \rightarrow H_3O^+(aq)$$
   In this reaction, water acts as a Lewis base because it...
   i. accepts protons
   ii. donates protons
   iii. accepts electrons
   iv. donates electrons
   B What is the concentration (in mol.dm^{-3}) of H_3O^+ ions in a NaOH solution which has a pH of 12 at 25^{0}C?
   i. 1×10^{12}
   ii. 1×10^{2}
   iii. 1×10^{-2}
   iv. 1×10^{-12}
   (IEB Paper 2, 2005)
   C When chlorine water is added to a solution of potassium bromide, bromine is produced. Which one of the following statements concerning this reaction is correct?
   i. Br^- is oxidised
   ii. Cl_2 is oxidised
   iii. Br^- is the oxidising agent
   iv. Cl^- is the oxidising agent
   (IEB Paper 2, 2005)
4. The stomach secretes gastric juice, which contains hydrochloric acid. The gastric juice helps with digestion. Sometimes there is an overproduction of acid, leading to heartburn or indigestion. Antacids, such as milk of magnesia, can be taken to neutralise the excess acid. Milk of magnesia is only slightly soluble in water and has the chemical formula $Mg(OH)_2$.

a Write a balanced chemical equation to show how the antacid reacts with the acid.

b The directions on the bottle recommend that children under the age of 12 years take one teaspoon of milk of magnesia, whereas adults can take two teaspoons of the antacid. Briefly explain why the dosages are different.

c Why is it not advisable to take an overdose of the antacid in the stomach? Refer to the hydrochloric acid concentration in the stomach in your answer.
In an experiment, 25.0 cm^3 of a standard solution of sodium carbonate of concentration 0.1 mol.dm^{-3} was used to neutralise 35.0 cm^3 of a solution of hydrochloric acid.

d Write a balanced chemical equation for the reaction.

e Calculate the concentration of the acid.

(DoE Grade 11 Exemplar, 2007)

Chapter 16

Reaction Rates - Grade 12

16.1 Introduction

Before we begin this section, it might be useful to think about some different types of reactions and how quickly or slowly they occur.

?

Exercise: Thinking about reaction rates
Think about each of the following reactions:

- rusting of metals
- photosynthesis
- weathering of rocks (e.g. limestone rocks being weathered by water)
- combustion

1. For each of the reactions above, write a chemical equation for the reaction that takes place.
2. How fast is each of these reactions? Rank them in order from the fastest to the slowest.
3. How did you decide which reaction was the fastest and which was the slowest?
4. Try to think of some other examples of chemical reactions. How fast or slow is each of these reactions, compared with those listed earlier?

In a chemical reaction, the substances that are undergoing the reaction are called the **reactants**, while the substances that form as a result of the reaction are called the **products**. The **reaction rate** describes how quickly or slowly the reaction takes place. So how do we know whether a reaction is slow or fast? One way of knowing is to look either at how quickly the *reactants are used up* during the reaction or at how quickly the *product forms*. For example, iron and sulfur react according to the following equation:

$$Fe + S \rightarrow FeS$$

In this reaction, we can see the speed of the reaction by observing how long it takes before there is no iron or sulfur left in the reaction vessel. In other words, the reactants have been used up. Alternatively, one could see how quickly the iron sulfide product forms. Since iron sulfide looks very different from either of its reactants, this is easy to do.

In another example:

$$2Mg(s) + O_2 \rightarrow 2MgO(s)$$

In this case, the reaction rate depends on the speed at which the reactants (solid magnesium and oxygen gas) are used up, or the speed at which the product (magnesium oxide) is formed.

Definition: Reaction rate
The *rate of a reaction* describes how quickly *reactants are used up* or how quickly *products are formed* during a chemical reaction. The units used are: moles per second (mols/second or mol.s^{-1}).

The average rate of a reaction is expressed as the number of moles of reactant used up, divided by the total reaction time, or as the number of moles of product formed, divided by the reaction time. Using the magnesium reaction shown earlier:

$$Average\ reaction\ rate = \frac{moles\ Mg\ used}{reaction\ time\ (s)}$$

or

$$Average\ reaction\ rate = \frac{moles\ O_2\ used}{reaction\ time\ (s)}$$

or

$$Average\ reaction\ rate = \frac{moles\ MgO\ produced}{reaction\ time\ (s)}$$

Worked Example 76: Reaction rates

Question: The following reaction takes place:

$$4Li + O_2 \rightarrow 2Li_2O$$

After two minutes , 4 g of Lithium has been used up. Calculate the rate of the reaction.

Answer
Step 1 : Calculate the number of moles of Lithium that are used up in the reaction.

$$n = \frac{m}{M} = \frac{4}{6.94} = 0.58mols$$

Step 2 : Calculate the time (in seconds) for the reaction.

$$t = 2 \times 60 = 120s$$

Step 3 : Calculate the rate of the reaction.
Rate of reaction =

$$\frac{moles\ of\ Lithium\ used}{time} = \frac{0.58}{120} = 0.005$$

The rate of the reaction is 0.005 mol.s^{-1}

Exercise: Reaction rates

1. A number of different reactions take place. The table below shows the number of moles of reactant that are used up in a particular time for each reaction.

Reaction	Mols used up	Time	Reaction rate
1	2	30 secs	
2	5	2 mins	
3	1	1.5 mins	
4	3.2	1.5 mins	
5	5.9	30 secs	

   (a) Complete the table by calculating the rate of each reaction.
   (b) Which is the *fastest* reaction?
   (c) Which is the *slowest* reaction?

2. Two reactions occur simultaneously in separate reaction vessels. The reactions are as follows:

$$Mg + Cl_2 \rightarrow MgCl_2$$
$$2Na + Cl_2 \rightarrow 2NaCl$$

   After 1 minute, 2 g of $MgCl_2$ have been produced in the first reaction.

   (a) How many moles of $MgCl_2$ are produced after 1 minute?
   (b) Calculate the rate of the reaction, using the amount of product that is produced.
   (c) Assuming that the second reaction also proceeds at the same rate, calculate...
      i. the number of moles of NaCl produced after 1 minute.
      ii. the mass (in g) of sodium that is needed for this reaction to take place.

16.2 Factors affecting reaction rates

Several factors affect the rate of a reaction. It is important to know these factors so that reaction rates can be controlled. This is particularly important when it comes to industrial reactions, so that productivity can be maximised. The following are some of the factors that affect the rate of a reaction.

1. **Nature of reactants**

   Substances have different chemical properties and therefore react differently and at different rates.

2. **Concentration** (or **pressure** in the case of gases)

   As the concentration of the reactants increases, so does the reaction rate.

3. **Temperature**

   If the temperature of the reaction increases, so does the rate of the reaction.

4. **Catalyst**

   Adding a catalyst increases the reaction rate.

5. **Surface area of solid reactants**

   Increasing the surface area of the reactants (e.g. if a solid reactant is finely broken up) will increase the reaction rate.

Activity :: Experiment : The nature of reactants.

Aim:

To determine the effect of the nature of reactants on the rate of a reaction.

Apparatus:

Oxalic acid ($(COOH)_2$), iron(II) sulphate ($FeSO_4$), potassium permanganate ($KMnO_4$), concentrated sulfuric acid (H_2SO_4), spatula, test tubes, medicine dropper, glass beaker and glass rod.

Method:

1. In the first test tube, prepare an iron (II) sulphate solution by dissolving about two spatula points of iron (II) sulphate in 10 cm^3 of water.
2. In the second test tube, prepare a solution of oxalic acid in the same way.
3. Prepare a solution of sulfuric acid by adding 1 cm^3 of the concentrated acid to about 4 cm^3 of water. Remember always to add the *acid to the water*, and never the other way around.
4. Add 2 cm^3 of the sulfuric acid solution to the iron(II) and oxalic acid solutions respectively.
5. Using the medicine dropper, add a few drops of potassium permanganate to the two test tubes. Once you have done this, observe how quickly each solution discolours the potassium permanganate solution.

Results:

- You should have seen that the oxalic acid solution discolours the potassium permanganate much more slowly than the iron(II) sulphate.
- It is the oxalate ions ($C_2O_4^{2-}$) and the Fe^{2+} ions that cause the discolouration. It is clear that the Fe^{2+} ions act much more quickly than the $C_2O_4^{2-}$ ions. The reason for this is that there are no covalent bonds to be broken in the ions before the reaction can take place. In the case of the oxalate ions, covalent bonds between carbon and oxygen atoms must be broken first.

Conclusions:

The nature of the reactants can affect the rate of a reaction.

Interesting Fact

Oxalic acids are abundant in many plants. The leaves of the tea plant (*Camellia sinensis*) contain very high concentrations of oxalic acid relative to other plants. Oxalic acid also occurs in small amounts in foods such as parsley, chocolate, nuts and berries. Oxalic acid irritates the lining of the gut when it is eaten, and can be fatal in very large doses.

Activity :: Experiment : Surface area and reaction rates.

Marble ($CaCO_3$) reacts with hydrochloric acid (HCl) to form calcium chloride, water and carbon dioxide gas according to the following equation:

$$CaCO_3 + 2HCl \rightarrow CaCl_2 + H_2O + CO_2$$

Aim:

To determine the effect of the surface area of reactants on the rate of the reaction.

Apparatus:

2 g marble chips, 2 g powdered marble, hydrochloric acid, beaker, two test tubes.

Method:

1. Prepare a solution of hydrochloric acid in the beaker by adding 2 cm^3 of the concentrated solution to 20 cm^3 of water.
2. Place the marble chips and powdered marble into separate test tubes.
3. Add 10 cm^3 of the dilute hydrochloric acid to each of the test tubes and observe the rate at which carbon dioxide gas is produced.

Results:

- Which reaction proceeds the fastest?
- Can you explain this?

Conclusions:

The reaction with powdered marble is the fastest. The smaller the pieces of marble are, the greater the surface area for the reaction to take place. The greater the surface area of the reactants, the faster the reaction rate will be.

Activity :: Experiment : Reactant concentration and reaction rate.

Aim:

To determine the effect of reactant concentration on reaction rate.

Apparatus:

Concentrated hydrochloric acid (HCl), magnesium ribbon, two beakers, two test tubes, measuring cylinder.

Method:

1. Prepare a solution of dilute hydrochloric acid in one of the beakers by diluting 1 part concentrated acid with 10 parts water. For example, if you measure 1 cm^3 of concentrated acid in a measuring cylinder and pour it into a beaker, you will need to add 10 cm^3 of water to the beaker as well. In the same way, if you pour 2 cm^3 of concentrated acid into a beaker, you will need to add 20 cm^3 of water. Both of these are 1:10 solutions. Pour 10 cm^3 of the 1:10 solution into a test tube and mark it 'A'. Remember to add the *acid* to the *water*, and not the other way around.
2. Prepare a second solution of dilute hydrochloric acid by diluting 1 part concentrated acid with 20 parts water. Pour 10cm^3 of this 1:20 solution into a second test tube and mark it 'B'.
3. Take two pieces of magnesium ribbon of the **same length**. At the same time, put one piece of magnesium ribbon into test tube A and the other into test tube B, and observe closely what happens.

The equation for the reaction is:

$$2HCl + Mg \rightarrow MgCl_2 + H_2$$

Results:

- Which of the two solutions is more concentrated, the 1:10 or 1:20 hydrochloric acid solution?
- In which of the test tubes is the reaction the fastest? Suggest a reason for this.
- How can you measure the rate of this reaction?
- What is the gas that is given off?
- Why was it important that the same length of magnesium ribbon was used for each reaction?

Conclusions:

The 1:10 solution is more concentrated and this reaction therefore proceeds faster. The greater the concentration of the reactants, the faster the rate of the reaction. The rate of the reaction can be measured by the rate at which hydrogen gas is produced.

Activity :: Group work : The effect of temperature on reaction rate

1. In groups of 4-6, design an experiment that will help you to see the effect of temperature on the reaction time of 2 cm of magnesium ribbon and 20 ml of vinegar. During your group discussion, you should think about the following:
   - What equipment will you need?
   - How will you conduct the experiment to make sure that you are able to compare the results for different temperatures?
   - How will you record your results?
   - What safety precautions will you need to take when you carry out this experiment?
2. Present your experiment ideas to the rest of the class, and give them a chance to comment on what you have done.
3. Once you have received feedback, carry out the experiment and record your results.
4. What can you conclude from your experiment?

16.3 Reaction rates and collision theory

It should be clear now that the rate of a reaction varies depending on a number of factors. But how can we *explain* why reactions take place at different speeds under different conditions? Why, for example, does an increase in the surface area of the reactants also increase the rate of the reaction? One way to explain this is to use **collision theory.**

For a reaction to occur, the particles that are reacting must collide with one another. Only a fraction of all the collisions that take place actually cause a chemical change. These are called 'successful' collisions. When there is an increase in the *concentration* of reactants, the chance that reactant particles will collide with each other also increases because there are more particles in that space. In other words, the *collision frequency* of the reactants increases. The number of *successful* collisions will therefore also increase, and so will the rate of the reaction. In the same way, if the *surface area* of the reactants increases, there is also a greater chance that successful collisions will occur.

Definition: Collision theory
Collision theory is a theory that explains how chemical reactions occur and why reaction rates differ for different reactions. The theory assumes that for a reaction to occur the reactant particles must collide, but that only a certain fraction of the total collisions, the *effective collisions*, actually cause the reactant molecules to change into products. This is because only a small number of the molecules have enough energy and the right orientation at the moment of impact to break the existing bonds and form new bonds.

When the *temperature* of the reaction increases, the average kinetic energy of the reactant particles increases and they will move around much more actively. They are therefore more likely to collide with one another (Figure 16.1). Increasing the temperature also increases the number of particles whose energy will be greater than the activation energy for the reaction (refer section 16.5).

Figure 16.1: An increase in the temperature of a reaction increases the chances that the reactant particles (A and B) will collide because the particles have more energy and move around more.

Exercise: Rates of reaction

Hydrochloric acid and calcium carbonate react according to the following equation:

$$CaCO_3 + 2HCl \rightarrow CaCl_2 + H_2O + CO_2$$

The volume of carbon dioxide that is produced during the reaction is measured at different times. The results are shown in the table below.

Time (mins)	Volume of CO_2 produced (cm^3)
1	14
2	26
3	36
4	44
5	50
6	58
7	65
8	70
9	74
10	77

Note: On a graph of production against time, it is the gradient of the graph that shows the rate of the reaction.

Questions:

1. Use the data in the table to draw a graph showing the volume of gas that is produced in the reaction, over a period of 10 minutes.
2. At which of the following times is the reaction *fastest*? Time = 1 minute; time = 6 minutes or time = 8 minutes.
3. Suggest a reason why the reaction slows down over time.
4. Use the graph to estimate the volume of gas that will have been produced after 11 minutes.
5. After what time do you think the reaction will stop?
6. If the experiment was repeated using a more concentrated hydrochloric acid solution...
   (a) would the rate of the reaction increase or decrease from the one shown in the graph?
   (b) draw a rough line on the graph to show how you would expect the reaction to proceed with a more concentrated HCl solution.

16.4 Measuring Rates of Reaction

How the rate of a reaction is measured will depend on what the reaction is, and what product forms. Look back to the reactions that have been discussed so far. In each case, how was the rate of the reaction measured? The following examples will give you some ideas about other ways to measure the rate of a reaction:

- *Reactions that produce hydrogen gas:*

  When a metal dissolves in an acid, hydrogen gas is produced. A lit splint can be used to test for hydrogen. The 'pop' sound shows that hydrogen is present. For example, magnesium reacts with sulfuric acid to produce magnesium sulphate and hydrogen.

$$Mg(s) + H_2SO_4 \rightarrow MgSO_4 + H_2$$

- *Reactions that produce carbon dioxide:*

  When a carbonate dissolves in an acid, carbon dioxide gas is produced. When carbon dioxide is passes through limewater, it turns the limewater milky. This is the test for the presence of carbon dioxide. For example, calcium carbonate reacts with hydrochloric acid to produce calcium chloride, water and carbon dioxide.

$$CaCO_3(s) + 2HCl(aq) \rightarrow CaCl_2(aq) + H_2O(l) + CO_2(g)$$

- *Reactions that produce gases such as oxygen or carbon dioxide:*

  Hydrogen peroxide decomposes to produce oxygen. The volume of oxygen produced can be measured using the gas syringe method (figure 16.2). The gas collects in the syringe, pushing out against the plunger. The volume of gas that has been produced can be read from the markings on the syringe. For example, hydrogen peroxide decomposes in the presence of a manganese(IV) oxide catalyst to produce oxygen and water.

$$2H_2O_2(aq) \rightarrow 2H_2O(l) + O_2(g)$$

Gas Syringe System

[glassType=erlen,niveauLiquide1=40,tubeCoude]

Gas

Reactants

Figure 16.2: Gas Syringe Method

- *Precipitate reactions:*

  In reactions where a *precipitate* is formed, the amount of precipitate formed in a period of time can be used as a measure of the reaction rate. For example, when sodium thiosulphate reacts with an acid, a yellow precipitate of sulfur is formed. The reaction is as follows:

$$Na_2S_2O_3(aq) + 2HCl(aq) \rightarrow 2NaCl(aq) + SO_2(aq) + H_2O(l) + S(s)$$

One way to estimate the rate of this reaction is to carry out the investigation in a conical flask and to place a piece of paper with a black cross underneath the bottom of the flask. At the beginning of the reaction, the cross will be clearly visible when you look into the flask (figure 16.3). However, as the reaction progresses and more precipitate is formed, the cross will gradually become less clear and will eventually disappear altogether. Noting the time that it takes for this to happen will give an idea of the reaction rate. Note that it is not possible to collect the SO_2 gas that is produced in the reaction, because it is very soluble in water.

[glassType=erlen,niveauLiquide1=40]

Figure 16.3: At the beginning of the reaction beteen sodium thiosulphate and hydrochloric acid, when no precipitate has been formed, the cross at the bottom of the conical flask can be clearly seen.

- *Changes in mass:*

  The rate of a reaction that produces a gas can also be measured by calculating the mass loss as the gas is formed and escapes from the reaction flask. This method can be used for reactions that produce carbon dioxide or oxygen, but are not very accurate for reactions that give off hydrogen because the mass is too low for accuracy. Measuring changes in mass may also be suitable for other types of reactions.

Activity :: Experiment : Measuring reaction rates

Aim:

To measure the effect of concentration on the rate of a reaction.

Apparatus:

- 300 cm^3 of sodium thiosulphate ($Na_2S_2O_3$) solution. Prepare a solution of sodium thiosulphate by adding 12 g of $Na_2S_2O_3$ to 300 cm^3 of water. This is solution 'A'.
- 300 cm^3 of water
- 100 cm^3 of 1:10 dilute hydrochloric acid. This is solution 'B'.
- Six 100 cm^3 glass beakers
- Measuring cylinders
- Paper and marking pen
- Stopwatch or timer

Method:

One way to measure the rate of this reaction is to place a piece of paper with a cross underneath the reaction beaker to see how quickly the cross is made invisible by the formation of the sulfur precipitate.

1. Set up six beakers on a flat surface and mark them from 1 to 6. Under each beaker you will need to place a piece of paper with a large black cross.
2. Pour 60 cm^3 solution A into the first beaker and add 20 cm^3 of water
3. Use the measuring cylinder to measure 10 cm^3 HCl. Now add this HCl to the solution that is already in the first beaker (NB: Make sure that you always clean out the measuring cylinder you have used before using it for another chemical).
4. Using a stopwatch with seconds, record the time it takes for the precipitate that forms to block out the cross.
5. Now measure 50 cm^3 of solution A into the second beaker and add 30 cm^3 of water. To this second beaker, add 10 cm^3 HCl, time the reaction and record the results as you did before.
6. Continue the experiment by diluting solution A as shown below.

Beaker	**Solution A (cm^3)**	**Water (cm^3)**	**Solution B (cm^3)**	**Time (s)**
1	60	20	10	
2	50	30	10	
3	40	40	10	
4	30	50	10	
5	20	60	10	
6	10	70	10	

The equation for the reaction between sodium thiosulphate and hydrochloric acid is:

$$Na_2S_2O_3(aq) + 2HCl(aq) \rightarrow 2NaCl(aq) + SO_2(aq) + H_2O(l) + S(s)$$

Results:

- Calculate the reaction rate in each beaker. This can be done using the following equation:

$$Rate\ of\ reaction = \frac{1}{time}$$

- Represent your results on a graph. **Concentration** will be on the x-axis and **reaction rate** on the y-axis. Note that the original volume of $Na_2S_2O_3$ can be used as a measure of concentration.
- Why was it important to keep the volume of HCl constant?
- Describe the relationship between concentration and reaction rate.

Conclusions:

The rate of the reaction is fastest when the concentration of the reactants was the highest.

16.5 Mechanism of reaction and catalysis

Earlier it was mentioned that it is the *collision* of particles that causes reactions to occur and that only some of these collisions are 'successful'. This is because the reactant particles have a wide range of kinetic energy, and only a small fraction of the particles will have enough energy to actually break bonds so that a chemical reaction can take place. The minimum energy that is needed for a reaction to take place is called the **activation energy**. For more information on the energy of reactions, refer to chapter 14.

Definition: Activation energy
The energy that is needed to break the bonds in reactant molecules so that a chemical reaction can proceed.

Even at a fixed temperature, the energy of the particles varies, meaning that only some of them will have enough energy to be part of the chemical reaction, depending on the activation energy for that reaction. This is shown in figure 16.4. Increasing the reaction temperature has the effect of increasing the number of particles with enough energy to take part in the reaction, and so the reaction rate increases.

Figure 16.4: The distribution of particle kinetic energies at a fixed temperature

A **catalyst** functions slightly differently. The function of a catalyst is to lower the activation energy so that more particles now have enough energy to react. The catalyst itself is not changed during the reaction, but simply provides an alternative pathway for the reaction, so that it needs less energy. Some *metals* e.g. platinum, copper and iron can act as catalysts in certain reactions. In our own human bodies, *enzymes* are catalysts that help to speed up biological reactions. Catalysts generally react with one or more of the reactants to form a chemical intermediate which then reacts to form the final product. The chemical intermediate is sometimes called the **activated complex**.

The following is an example of how a reaction that involves a catalyst might proceed. C represents the catalyst, A and B are reactants and D is the product of the reaction of A and B.

Step 1: A + C → AC

Step 2: B + AC → ABC

Step 3: ABC → CD

Step 4: CD → C + D

In the above, ABC represents the intermediate chemical. Although the catalyst (C) is consumed by reaction 1, it is later produced again by reaction 4, so that the overall reaction is as follows:

$$A + B + C \rightarrow D + C$$

You can see from this that the catalyst is released at the end of the reaction, completely unchanged.

Definition: Catalyst
A catalyst speeds up a chemical reaction, without being altered in any way. It increases the reaction rate by lowering the activation energy for a reaction.

Energy diagrams are useful to illustrate the effect of a **catalyst** on reaction rates. Catalysts decrease the activation energy required for a reaction to proceed (shown by the smaller 'hump' on the energy diagram in figure 16.5), and therefore increase the reaction rate.

Figure 16.5: The effect of a catalyst on the activation energy of a reaction

---

Activity :: Experiment : Catalysts and reaction rates

Aim:

To determine the effect of a catalyst on the rate of a reaction

Apparatus:

Zinc granules, 0.1 M hydrochloric acid, copper pieces, one test tube and a glass beaker.

Method:

1. Place a few of the zinc granules in the test tube.
2. Measure the mass of a few pieces of copper and keep them separate from the rest of the copper.
3. Add about 20 cm^3 of HCl to the test tube. You will see that a gas is released. Take note of how quickly or slowly this gas is released. Write a balanced equation for the chemical reaction that takes place.
4. Now add the copper pieces to the same test tube. What happens to the rate at which the gas is produced?
5. Carefully remove the copper pieces from the test tube (do not get HCl on your hands), rinse them in water and alcohol and then weigh them again. Has the mass of the copper changed since the start of the experiment?

Results:

During the reaction, the gas that is released is hydrogen. The rate at which the hydrogen is produced *increases* when the copper pieces (the catalyst) are added. The mass of the copper does not change during the reaction.

Conclusions:

The copper acts as a *catalyst* during the reaction. It speeds up the rate of the reaction, but is not changed in any way itself.

Exercise: Reaction rates

1. For each of the following, say whether the statement is **true** or **false**. If it is false, re-write the statement correctly.
   (a) A catalyst increases the energy of reactant molecules so that a chemical reaction can take place.
   (b) Increasing the temperature of a reaction has the effect of increasing the number of reactant particles that have more energy that the activation energy.
   (c) A catalyst does not become part of the final product in a chemical reaction.
2. 5 g of zinc granules are added to 400 cm^3 of 0.5 $mol.dm^{-3}$ hydrochloric acid. To investigate the rate of the reaction, the change in the mass of the flask containing the zinc and the acid was measured by placing the flask on a direct reading balance. The reading on the balance shows that there is a decrease in mass during the reaction. The reaction which takes place is given by the following equation:

$$Zn(s) + 2HCl(aq) \rightarrow ZnCl_2(aq) + H_2(g)$$

   (a) Why is there a decrease in mass during the reaction?
   (b) The experiment is repeated, this time using 5 g of powdered zinc instead of granulated zinc. How will this influence the rate of the reaction?
   (c) The experiment is repeated once more, this time using 5 g of granulated zinc and 600 cm^3 of 0.5 $mol.dm^{-3}$ hydrochloric acid. How does the rate of this reaction compare to the **original** reaction rate?
   (d) What effect would a catalyst have on the rate of this reaction?

   (IEB Paper 2 2003)
3. Enzymes are catalysts. Conduct your own research to find the names of common enzymes in the human body and which chemical reactions they play a role in.
4. 5 g of calcium carbonate powder reacts with 20 cm^3 of a 0.1 $mol.dm^{-3}$ solution of hydrochloric acid. The gas that is produced at a temperature of 25^0C is collected in a gas syringe.
   (a) Write a balanced chemical equation for this reaction.
   (b) The rate of the reaction is determined by measuring the volume of gas thas is produced in the first minute of the reaction. How would the rate of the reaction be affected if:
      i. a lump of calcium carbonate of the same mass is used
      ii. 40 cm^3 of 0.1 $mol.dm^{-3}$ hydrochloric acid is used

16.6 Chemical equilibrium

Having looked at factors that affect the rate of a reaction, we now need to ask some important questions. Does a reaction always proceed in the same direction or can it be reversible? In other words, is it always true that a reaction proceeds from *reactants to products*, or is it possible that sometimes, the reaction will reverse and the *products will be changed back into the reactants*? And does a reaction always run its full course so that all the reactants are used up, or can a reaction reach a point where reactants are still present, but there does not seem to be any further change taking place in the reaction? The following demonstration might help to explain this.

Activity :: Demonstration : Liquid-vapour phase equilibrium
Apparatus and materials:
2 beakers; water; bell jar
Method:

1. Half fill two beakers with water and mark the level of the water in each case.
2. Cover one of the beakers with a bell jar.
3. Leave the beakers and, over the course of a day or two, observe how the water level in the two beakers changes. What do you notice? Note: You could speed up this demonstration by placing the two beakers over a bunsen burner to heat the water. In this case, it may be easier to cover the second beaker with a glass cover.

Observations:
You should notice that in the beaker that is uncovered, the water level drops quickly because of evaporation. In the beaker that is covered, there is an initial drop in the water level, but after a while evaporation appears to stop and the water level in this beaker is higher than that in the one that is open. Note that the diagram below shows the situation ate time=0.

Discussion:
In the first beaker, liquid water becomes water vapour as a result of evaporation and the water level drops. In the second beaker, evaporation also takes place. However, in this case, the vapour comes into contact with the surface of the bell jar and it cools and condenses to form liquid water again. This water is returned to the beaker. Once condensation has begun, the rate at which water is lost from the beaker will start to decrease. At some point, the rate of evaporation will be equal to the rate of condensation above the beaker, and there will be no change in the water level in the beaker. This can be represented as follows:

$$liquid \Leftrightarrow vapour$$

In this example, the reaction (in this case, a change in the phase of water) can proceed in either direction. In one direction there is a change in phase from liquid to vapour. But the reverse can also take place, when vapour condenses to form water again.

In a **closed system** it is possible for reactions to be reversible, such as in the demonstration above. In a closed system, it is also possible for a chemical reaction to reach **equilibrium**. We will discuss these concepts in more detail.

16.6.1 Open and closed systems

An **open system** is one in which matter or energy can flow into or out of the system. In the liquid-vapour demonstration we used, the first beaker was an example of an open system because the beaker could be heated or cooled (a change in *energy*), and water vapour (the *matter*) could evaporate from the beaker.

A **closed system** is one in which energy can enter or leave, but matter cannot. The second beaker covered by the bell jar is an example of a closed system. The beaker can still be heated or cooled, but water vapour cannot leave the system because the bell jar is a barrier. Condensation changes the vapour to liquid and returns it to the beaker. In other words, there is no loss of matter from the system.

Definition: Open and closed systems
An *open system* is one whose borders allow the movement of energy and matter into and out of the system. A *closed system* is one in which only energy can be exchanged, but not matter.

16.6.2 Reversible reactions

Some reactions can take place in two directions. In one direction the reactants combine to form the products. This is called the **forward reaction**. In the other, the products react to form reactants again. This is called the **reverse reaction**. A special double-headed arrow is used to show this type of **reversible reaction**:

$$XY + Z \Leftrightarrow X + YZ$$

So, in the following reversible reaction:

$$H_2(g) + I_2(g) \Leftrightarrow 2HI(g)$$

The forward reaction is $H_2(g) + I_2(g) \rightarrow 2HI(g)$. The reverse reaction is $2HI(g) \rightarrow H_2(g) + I_2(g)$.

Definition: A reversible reaction
A reversible reaction is a chemical reaction that can proceed in both the forward and reverse directions. In other words, the reactant and product of one reaction may reverse roles.

Activity :: Demonstration : The reversibility of chemical reactions
Apparatus and materials:
Lime water ($Ca(OH)_2$); calcium carbonate ($CaCO_3$); hydrochloric acid; 2 test tubes with rubber stoppers; delivery tube; retort stand and clamp; bunsen burner.
Method and observations:

1. Half-fill a test tube with clear lime water ($Ca(OH)_2$).
2. In another test tube, place a few pieces of calcium carbonate ($CaCO_3$) and cover the pieces with dilute hydrochloric acid. Seal the test tube with a rubber stopper and delivery tube.
3. Place the other end of the delivery tube into the test tube containing the lime water so that the carbon dioxide that is produced from the reaction between calcium carbonate and hydrochloric acid passes through the lime water. Observe what happens to the appearance of the lime water.

   The equation for the reaction that takes place is:

$$Ca(OH)_2 + CO_2 \rightarrow CaCO_3 + H_2O$$

$CaCO_3$ is insoluble and it turns the limewater milky.

4. Allow the reaction to proceed for a while so that carbon dioxide continues to pass through the limewater. What do you notice? The equation for the reaction that takes place is:

$$CaCO_3(s) + H_2O + CO_2 \rightarrow Ca(HCO_3)_2$$

In this reaction, calcium carbonate becomes one of the reactants to produce hydrogen carbonate ($Ca(HCO_3)_2$) and so the solution becomes clear again.

5. Heat the solution in the test tube over a bunsen burner. What do you observe? You should see bubbles of carbon dioxide appear and the limewater turns milky again. The reaction that has taken place is:

$$Ca(HCO_3)_2 \rightarrow CaCO_3(s) + H_2O + CO_2$$

Discussion:

- If you look at the last two equations you will see that the one is the reverse of the other. In other words, this is a *reversible reaction* and can be written as follows:

$$CaCO_3(s) + H_2O + CO_2 \Leftrightarrow Ca(HCO_3)_2$$

- Is the forward reaction endothermic or exothermic? Is the reverse reaction endothermic or exothermic? You should have noticed that the reverse reaction only took place when the solution was heated. Sometimes, changing the temperature of a reaction can change its direction.

16.6.3 Chemical equilibrium

Using the same reversible reaction that we used in an earlier example:

$$H_2(g) + I_2(g) \Leftrightarrow 2HI(g)$$

The forward reaction is:

$$H_2 + I_2 \rightarrow 2HI$$

The reverse reaction is:

$$2HI \rightarrow H_2 + I_2$$

When the rate of the forward reaction and the reverse reaction are equal, the system is said to be in **equilbrium**. Figure 16.6 shows this. Initially (time = 0), the rate of the forward reaction is high and the rate of the reverse reaction is low. As the reaction proceeds, the rate of the forward reaction decreases and the rate of the reverse reaction increases, until both occur at the same rate. This is called equilibrium.

Although it is not always possible to observe any macroscopic changes, this does not mean that the reaction has stopped. The forward and reverse reactions continue to take place and so microscopic changes still occur in the system. This state is called **dynamic equilibrium**. In the liquid-vapour phase equilibrium demonstration, dynamic equilibrium was reached when there was no observable change in the level of the water in the second beaker even though evaporation and condensation continued to take place.

Figure 16.6: The change in rate of forward and reverse reactions in a closed system

There are, however, a number of factors that can change the chemical equilibrium of a reaction. Changing the **concentration**, the **temperature** or the **pressure** of a reaction can affect equilibrium. These factors will be discussed in more detail later in this chapter.

Definition: Chemical equilibrium
Chemical equilibrium is the state of a chemical reaction, where the concentrations of the reactants and products have no net change over time. Usually, this state results when the forward chemical reactions proceed at the same rate as their reverse reactions.

16.7 The equilibrium constant

Definition: Equilibrium constant
The equilibrium constant (K_c), relates to a chemical reaction at equilibrium. It can be calculated if the equilibrium concentration of each reactant and product in a reaction at equilibrium is known.

16.7.1 Calculating the equilibrium constant

Consider the following generalised reaction which takes place in a **closed** container at a **constant temperature**:

$$A + B \Leftrightarrow C + D$$

We know from section 16.2 that the rate of the forward reaction is directly proportional to the concentration of the reactants. In other words, as the concentration of the reactants increases, so does the rate of the forward reaction. This can be shown using the following equation:

Rate of forward reaction $\propto$ [A][B]

or

Rate of forward reaction = k_1[A][B]

Similarly, the rate of the reverse reaction is directly proportional to the concentration of the products. This can be shown using the following equation:

Rate of reverse reaction $\propto$ [C][D]

or

Rate of reverse reaction = k_2[C][D]

At equilibrium, the rate of the forward reaction is equal to the rate of the reverse reaction. This can be shown using the following equation:

$$k_1[A][B] = k_2[C][D]$$

or

$$\frac{k_1}{k_2} = \frac{[C][D]}{[A][B]}$$

or, if the constants k_1 and k_2 are simplified to a single constant, the equation becomes:

$$k_c = \frac{[C][D]}{[A][B]}$$

A more general form of the equation for a reaction at chemical equilibrium is:

$$aA + bB \Leftrightarrow cC + dD$$

where A and B are reactants, C and D are products and a, b, c, and d are the coefficients of the respective reactants and products. A more general formula for calculating the equilibrium constant is therefore:

$$k_c = \frac{[C]^c[D]^d}{[A]^a[B]^b}$$

It is important to note that if a reactant or a product in a chemical reaction is in either the **liquid** or **solid** phase, the concentration stays constant during the reaction. Therefore, these values can be left out of the equation to calculate k_c. For example, in the following reaction:

$$C(s) + H_2O(g) \Leftrightarrow CO(g) + H_2(g)$$

$$k_c = \frac{[CO][H_2]}{[H_2O]}$$

Important:

1. The constant k_c is affected by temperature and so, if the values of k^c are being compared for different reactions, it is important that all the reactions have taken place at the same temperature.
2. k_c values do not have units. If you look at the equation, the units all cancel each other out.

16.7.2 The meaning of k_c values

The formula for k_c has the concentration of the products in the numerator and the concentration of reactants in the denominator. So a *high* k_c value means that the concentration of products is high and the reaction has a high yield. We can also say that the equilibrium lies far to the right. The opposite is true for a low k_c value. A *low* k_c *value* means that, at equilibrium, there are more reactants than products and therefore the yield is low. The equilibrium for the reaction lies far to the left.

Important: Calculations made easy

When you are busy with calculations that involve the equilibrium constant, the following tips may help:

1. Make sure that you always read the question carefully to be sure of what you are being asked to calculate. If the equilibrium constant is involved, make sure that the concentrations you use are the concentrations **at equilibrium**, and not the concentrations or quantities that are present at some other time in the reaction.
2. When you are doing more complicated calculations, it sometimes helps to draw up a table like the one below and fill in the **mole values** that you know or those you can calculate. This will give you a clear picture of what is happening in the reaction and will make sure that you use the right values in your calculations.

	Reactant 1	**Reactant 2**	**Product 1**
Start of reaction			
Used up			
Produced			
Equilibrium			

Worked Example 77: Calculating k_c

Question: For the reaction:

$$SO_2(g) + NO_2(g) \rightarrow NO(g) + SO_3(g)$$

the concentration of the reagents is as follows:
[SO_3] = 0.2 mol.dm^{-3}
[NO_2] = 0.1 mol.dm^{-3}
[NO] = 0.4 mol.dm^{-3}
[SO_2] = 0.2 mol.dm^{-3}

Calculate the value of k_c.

Answer
Step 1 : Write the equation for k_c

$$k_c = \frac{[NO][SO_3]}{[SO_2][NO_2]}$$

Step 2 : Fill in the values you know for this equation and calculate k_c

$$k_c = \frac{(0.4 \times 0.2)}{(0.2 \times 0.1)} = 4$$

Worked Example 78: Calculating reagent concentration

Question: For the reaction:

$$S(s) + O_2(g) \Leftrightarrow SO_2(g)$$

1. Write an equation for the equilibrium constant.
2. Calculate the equilibrium concentration of O_2 if Kc=6 and $[SO_2] = 3mol.dm^{-3}$ at equilibrium.

Answer
Step 1 : Write the equation for k_c

$$k_c = \frac{[SO_2]}{[O_2]}$$

(Sulfur is left out of the equation because it is a solid and its concentration stays constant during the reaction)

Step 2 : Re-arrange the equation so that oxygen is on its own on one side of the equation

$$[O_2] = \frac{[SO_2]}{k_c}$$

Step 3 : Fill in the values you know and calculate [O_2]

$$[O_2] = \frac{3mol.dm^{-3}}{6} = 0.5mol.dm^{-3}$$

Worked Example 79: Equilibrium calculations

Question: Initially 1.4 moles of NH_3(g) is introduced into a sealed 2.0 dm^{-3} reaction vessel. The ammonia decomposes when the temperature is increased to 600K and reaches equilibrium as follows:

$$2NH_3(g) \Leftrightarrow N_2(g) + 3H_2(g)$$

When the equilibrium mixture is analysed, the concentration of NH_3(g) is 0.3 mol.dm^{-3}

1. Calculate the concentration of N_2(g) and H_2(g) in the equilibrium mixture.
2. Calculate the equilibrium constant for the reaction at 900 K.

Answer
Step 1 : Calculate the number of moles of NH_3 at equilibrium.

$$c = \frac{n}{V}$$

Therefore,

$$n = c \times V = 0.3 \times 2 = 0.6mol$$

Step 2 : Calculate the number of moles of ammonia that react (are 'used up') in the reaction.
Moles used up = 1.4 - 0.6 = 0.8 moles

Step 3 : Calculate the number of moles of product that are formed.
Remember to use the mole ratio of reactants to products to do this. In this case, the ratio of NH_3:N_2:H_2 = 2:1:3. Therefore, if 0.8 moles of ammonia are used up in the reaction, then 0.4 moles of nitrogen are produced and 1.2 moles of hydrogen are produced.

Step 4 : Complete the following table

	NH_3	**N_2**	**H_2**
Start of reaction	1.4	0	0
Used up	0.8	0	0
Produced	0	0.4	1.2
Equilibrium	0.6	0.4	1.2

Step 5 : Using the values in the table, calculate [N_2] and [H_2]

$$[N_2] = \frac{n}{V} = \frac{0.4}{2} = 0.2\ mol.dm^{-3}$$

$$[H_2] = \frac{n}{V} = \frac{1.2}{2} = 0.6\ mol.dm^{-3}$$

Step 6 : Calculate k$_c$

$$k_c = \frac{[H_2]^3[N_2]}{[NH_3]^2} = \frac{(0.6)^3(0.2)}{(0.3)^2} = 0.48$$

Worked Example 80: Calculating k$_c$

Question: Hydrogen and iodine gas react according to the following equation:

$H_2(g) + I_2(g) \Leftrightarrow 2HI(g)$

When 0.496 mol H_2 and 0.181 mol I_2 are heated at 450^oC in a 1 dm^3 container, the equilibrium mixture is found to contain 0.00749 mol I_2. Calculate the equilibrium constant for the reaction at 450^oC.

Answer
Step 1 : Calculate the number of moles of iodine used in the reaction.
Moles of iodine used = 0.181 - 0.00749 = 0.1735 mol

Step 2 : Calculate the number of moles of hydrogen that are used up in the reaction.
The mole ratio of hydrogen:iodine = 1:1, therefore 0.1735 moles of hydrogen must also be used up in the reaction.

Step 3 : Calculate the number of moles of hydrogen iodide that are produced.
The mole ratio of H_2:I_2:HI = 1:1:2, therefore the number of moles of HI produced is 0.1735 × 2 = 0.347 mol.

So far, the table can be filled in as follows:

	H_2 (g)	**I_2**	**2HI**
Start of reaction	0.496	0.181	0
Used up	0.1735	0.1735	0
Produced	0	0	0.347
Equilibrium	0.3225	0.0075	0.347

Step 4 : Calculate the concentration of each of the reactants and products at equilibrium.

$$c = \frac{n}{V}$$

Therefore the equilibrium concentrations are as follows:
[H_2] = 0.3225 mol.dm^{-3}
[I_2] = 0.0075 mol.dm^{-3}
[HI] = 0.347 mol.dm^{-3}

Step 5 : Calculate k_c

$$k_c = \frac{[HI]}{[H_2][I_2]} = \frac{0.347}{0.3225 \times 0.0075} = 143.47$$

Exercise: The equilibrium constant

1. Write the equilibrium constant expression, K_c for the following reactions:
   (a) $2\text{NO(g)} + \text{Cl}_2\text{(g)} \Leftrightarrow 2\text{NOCl}$
   (b) $\text{H}_2\text{(g)} + \text{I}_2\text{(g)} \Leftrightarrow 2\text{HI(g)}$
2. The following reaction takes place:
   $$\text{Fe}^{3+}\text{(aq)} + 4\text{Cl}^- \Leftrightarrow \text{FeCl}_4^-\text{(aq)}$$
   K_c for the reaction is 7.5 × 10^{-2} mol.dm^{-3}. At equilibrium, the concentration of $FeCl_4^-$ is 0.95 × 10^{-4} mol.dm^{-3} and the concentration of free iron (Fe^{3+}) is 0.2 mol.dm^{-3}. Calculate the concentration of chloride ions at equilibrium.
3. Ethanoic acid (CH_3COOH) reacts with ethanol (CH_3CH_2OH) to produce ethyl ethanoate and water. The reaction is:
   $$\text{CH}_3\text{COOH} + \text{CH}_3\text{CH}_2\text{OH} \rightarrow \text{CH}_3\text{COOCH}_2\text{CH}_3 + \text{H}_2\text{O}$$
   At the beginning of the reaction, there are 0.5 mols of ethanoic acid and 0.5 mols of ethanol. At equilibrium, 0.3 mols of ethanoic acid was left unreacted. The volume of the reaction container is 2 dm^3. Calculate the value of K_c.

16.8 Le Chatelier's principle

A number of factors can influence the equilibrium of a reaction. These are:

1. concentration
2. temperature
3. pressure

Le Chatelier's Principle helps to predict what a change in temperature, concentration or pressure will have on the position of the equilibrium in a chemical reaction. This is very important, particularly in industrial applications, where yields must be accurately predicted and maximised.

Definition: Le Chatelier's Principle
If a chemical system at equilibrium experiences a change in concentration, temperature or total pressure the equilibrium will shift in order to minimise that change.

16.8.1 The effect of concentration on equilibrium

If the concentration of a substance is increased, the equilibrium will shift so that this concentration decreases. So for example, if the concentration of a reactant was increased, the equilibrium would shift in the direction of the reaction that *uses up* the reactants, so that the reactant concentration decreases and equilibrium is restored. In the reaction between nitrogen and hydrogen to produce ammonia:

$$N_2(g) + 3H_2(g) \Leftrightarrow 2NH_3(g)$$

- If the nitrogen or hydrogen concentration was increased, Le Chatelier's principle predicts that equilibrium will shift to favour the *forward reaction* so that the excess nitrogen and hydrogen are used up to produce ammonia. Equilibrium shifts to the *right*.
- If the nitrogen or hydrogen concentration was decreased, the *reverse reaction* would be favoured so that some of the ammonia would change back to nitrogen and hydrogen to restore equilibrium.
- The same would be true if the concentration of the product (NH_3) was changed. If [NH_3] decreases, the forward reaction is favoured and if [NH_3] increases, the reverse reaction is favoured.

16.8.2 The effect of temperature on equilibrium

If the temperature of a reaction mixture is increased, the equilibrium will shift to decrease the temperature. So it will favour the reaction which will *use up* heat energy, in other words the endothermic reaction. The opposite is true if the temperature is decreased. In this case, the reaction that *produces* heat energy will be favoured, in other words, the exothermic reaction.

The reaction shown below is exothermic (shown by the negative value for Δ H). This means that the forward reaction, where nitrogen and hydrogen react to form ammonia, gives off heat. In the reverse reaction, where ammonia is broken down into hydrogen and nitrogen gas, heat is used up and so this reaction is endothermic.

$$e.g. \quad N_2(g) + 3H_2(g) \Leftrightarrow 2NH_3(g) \text{ and } \Delta\text{H} = -92\text{kJ}$$

An increase in temperature favours the reaction that is endothermic (the reverse reaction) because it uses up energy. If the temperature is increased, then the yield of ammonia (NH_3)

decreases.

A decrease in temperature favours the reaction that is exothermic (the forward reaction) because it produces energy. Therefore, if the temperature is decreased, then the yield of NH_3 increases.

Activity :: Experiment : Le Chatelier's Principle

Aim:

To determine the effect of a change in concentration and temperature on chemical equilibrium

Apparatus:

0.2 M $CoCl_2$ solution, concentrated HCl, water, test tube, bunsen burner

Method:

1. Put 4-5 drops of 0.2M $CoCl_2$ solution into a test tube.
2. Add 20-25 drops of concentrated HCl.
3. Add 10-12 drops of water.
4. Heat the solution for 1-2 minutes.
5. Cool the solution for 1 minute under a tap.
6. Observe and record the colour changes that take place during the reaction.

The equation for the reaction that takes place is:

$$e.g. \quad \underbrace{CoCl_4^{2-} + 6H_2O}_{\text{blue}} \Leftrightarrow \underbrace{Co(H_2O)_6^{2+} + 4Cl^-}_{\text{pink}}$$

Results:

Complete your observations in the table below, showing the colour changes that take place, and also indicating whether the concentration of each of the ions in solution increases or decreases.

	Initial colour	**Final colour**	**[Co^{2+}]**	**[Cl^-]**	**[$CoCl_4^{2-}$]**
Add Cl^-					
Add H_2O					
Increase temp.					
Decrease temp.					

Conclusions:

Use your knowledge of equilibrium principles to explain the changes that you recorded in the table above. Draw a conclusion about the effect of a change in concentration of either the reactants or products on the equilibrium position. Also draw a conclusion about the effect of a change in temperature on the equilibrium position.

16.8.3 The effect of pressure on equilibrium

In the case of gases, we refer to pressure instead of concentration. Similar principles apply as those that were described before for concentration. When the pressure of a system increases, there are more particles in a particular space. The equilibrium will shift in a direction that reduces the number of gas particles so that the pressure is also reduced. To predict what will happen in a reaction, we need to look at the number of moles of gas that are in the reactants and products. Look at the example below:

$$e.g. \quad 2SO_2(g) + O_2(g) \Leftrightarrow 2SO_3(g)$$

In this reaction, two moles of product are formed for every three moles of reactants. If we increase the pressure on the closed system, the equilibrium will shift to the right because the forward reaction reduces the number of moles of gas that are present. This means that the yield of SO_3 will increase. The opposite will apply if the pressure on the system decreases. the equilibrium will shift to the left, and the concentration of SO_2 and O_2 will increase.

Important: The following rules will help in predicting the changes that take place in equilibrium reactions:

1. If the forward reaction that forms the product is endothermic, then an increase in temperature will favour this reaction and the yield of product will increase. Lowering the temperature will decrease the product yield.
2. If the forward reaction that forms the product is exothermic, then a decrease in temperature will favour this reaction and the product yield will increase. Increasing the temperature will decrease the product yield.
3. Increasing the pressure favours the side of the equilibrium with the least number of gas molecules. This is shown in the balanced symbol equation. This rule applies in reactions with one or more gaseous reactants or products.
4. Decreasing the pressure favours the side of the equilibrium with the most number of gas molecules. This rule applies in reactions with one or more gaseous reactants or products.
5. If the concentration of a reactant (on the left) is increased, then some of it must change to the products (on the right) for equilibrium to be maintained. The equilibrium position will shift to the right.
6. If the concentration of a reactant (on the left) is decreased, then some of the products (on the right) must change back to reactants for equilibrium to be maintained. The equilibrium position will shift to the left.
7. A **catalyst** does not affect the equilibrium position of a reaction. It only influences the *rate of the reaction*, in other words, how quickly equilibrium is reached.

Worked Example 81: Reaction Rates 1

Question: $2NO_2(g) \Leftrightarrow 2NO(g) + O_2(g)$ and $\Delta H > 0$ How will the rate of the reverse reaction be affected by:

1. a decrease in temperature?
2. the addition of a catalyst?
3. the addition of more NO gas?

Answer

1. The rate of the forward reaction will increase since it is the forward reaction that is exothermix and therefore *produces* energy to balance the loss of energy from the decrease in temperature. The rate of the reverse reaction will decrease.
2. The rate of the reverse and the forward reaction will increase.
3. The rate of the reverse reaction will increase so that the extra NO gas is converted into NO_2 gas.

Worked Example 82: Reaction Rates 2

Question:

1. Write a balanced equation for the exothermic reaction between Zn(s) and HCl.
2. Name 3 ways to increase the reaction rate between hydrochloric acid and zinc metal.

Answer

1. $Zn(s) + 2HCl(aq) \Leftrightarrow ZnCl_2(aq) + H_2(g)$
2. A catalyst could be added, the zinc solid could be ground into a fine powder to increase its surface area, the HCl concentration could be increased or the reaction temperature could be increased.

Exercise: Reaction rates and equilibrium

1. The following reaction reaches equilibrium in a closed container:

$$CaCO_3(s) \Leftrightarrow CaO(s) + CO_2(g)$$

The pressure of the system is increased by decreasing the volume of the container. How will the number of moles and the concentration of the CO_2(g) have changed when a new equilibrium is reached at the same temperature?

	moles of CO_2	**[CO_2]**
A	decreased	decreased
B	increased	increased
C	decreased	stays the same
D	decreased	increased

(IEB Paper 2, 2003)

2. The following reaction has reached equilibrium in a closed container:

$$C(s) + H_2O(g) \Leftrightarrow CO(g) + H_2(g) \ \Delta H \text{ ¿ } 0$$

The pressure of the system is then decreased by increasing the volume of the container. How will the concentration of the H_2(g) and the value of K_c be affected when the new equilibrium is established? Assume that the temperature of the system remains unchanged.

	[H_2]	**K_c**
A	increases	increases
B	increases	unchanged
C	unchanged	unchanged
D	decreases	unchanged

(IEB Paper 2, 2004)

3. During a classroom experiment copper metal reacts with concentrated nitric acid to produce NO_2 gas, which is collected in a gas syringe. When enough gas has collected in the syringe, the delivery tube is clamped so that no gas can escape. The brown NO_2 gas collected reaches an equilibrium with colourless N_2O_4 gas as represented by the following equation:

$$2NO_2(g) \Leftrightarrow N_2O_4(g)$$

Once this equilibrium has been established, there are 0.01 moles of NO_2 gas and 0.03 moles of N_2O_4 gas present in the syringe.

(a) A learner, noticing that the colour of the gas mixture in the syringe is no longer changing, comments that all chemical reactions in the syringe must have stopped. Is this assumption correct? Explain.

(b) The gas in the syringe is cooled. The volume of the gas is kept constant during the cooling process. Will the gas be lighter or darker at the lower temperature? Explain your answer.

(c) The volume of the syringe is now reduced to 75 cm^3 by pushing the plunger in and holding it in the new position. There are 0.032 moles of N_2O_4 gas present once the equilibrium has been re-established at the reduced volume (75 cm^3). Calculate the value of the equilibrium constant for this equilibrium.
(IEB Paper 2, 2004)

4. Consider the following reaction, which takes place in a closed container:

$$\mathrm{A(s) + B(g) \rightarrow AB(g)}\ \Delta H < 0$$

If you wanted to increase the rate of the reaction, which of the following would you do?

(a) decrease the concentration of B
(b) decrease the temperature of A
(c) grind A into a fine powder
(d) decrease the pressure

(IEB Paper 2, 2002)

5. Gases X and Y are pumped into a 2 dm^3 container. When the container is sealed, 4 moles of gas X and 4 moles of gas Y are present. The following equilibrium is established:

$$\mathrm{2X(g) + 3Y(g) \Leftrightarrow X_2Y_3}$$

The graph below shows the number of moles of gas X and gas X_2Y_3 that are present from the time the container is sealed.

(a) How many moles of gas X_2Y_3 are formed by the time the reaction reaches equilibrium at 30 seconds?

(b) Calculate the value of the equilibrium constant at t = 50 s.

(c) At 70 s the temperature is increased. Is the forward reaction endothermic or exothermic? Explain in terms of Le Chatelier's Principle.

(d) How will this increase in temperature affect the value of the equilibrium constant?

16.9 Industrial applications

The **Haber process** is a good example of an industrial process which uses the equilibrium principles that have been discussed. The equation for the process is as follows:

$$N_2(g) + 3H_2(g) \Leftrightarrow 2NH_3(g) + energy$$

Since the reaction is **exothermic**, the forward reaction is favoured at low temperatures, and the reverse reaction at high temperatures. If the purpose of the Haber process is to produce ammonia, then the temperature must be maintained at a level that is low enough to ensure that the reaction continues in the forward direction.

The forward reaction is also favoured by high *pressures* because there are four moles of reactant for every two moles of product formed.

The k value for this reaction will be calculated as follows:

$$k = \frac{[NH_3]^2}{[N_2][H_2]^3}$$

Exercise: Applying equilibrium principles

Look at the values of k calculated for the Haber process reaction at different temperatures, and then answer the questions that follow:

T^{oC}	k
25	6.4 x 10^2
200	4.4 x 10^{-1}
300	4.3 x 10^{-3}
400	1.6 x 10^{-4}
500	1.5 x 10^{-5}

1. What happens to the value of k as the temperature increases?
2. Which reaction is being favoured when the temperature is 300 degrees celsius?
3. According to this table, which temperature would be best if you wanted to produce as much ammonia as possible? Explain.

16.10 Summary

- The **rate of a reaction** describes how quickly reactants are used up, or how quickly products form. The units used are moles per second.
- A number of factors can affect the rate of a reaction. These include the **nature of the reactants**, the **concentration** of reactants, **temperature** of the reaction, the presence or absence of a **catalyst** and the **surface area** of the reactants.
- **Collision theory** provides one way of explaining why each of these factors can affect the rate of a reaction. For example, higher temperatures mean increased reaction rates because the reactant particles have more energy and are more likely to collide successfully with each other.
- Different methods can be used to **measure the rate of a reaction**. The method used will depend on the nature of the product. Reactions that produce gases can be measured by collecting the gas in a syringe. Reactions that produce a precipitate are also easy to measure because the precipitate is easily visible.
- For any reaction to occur, a minimum amount of energy is needed so that bonds in the reactants can break, and new bonds can form in the products. The minimum energy that is required is called the **activation energy** of a reaction.
- In reactions where the particles do not have enough energy to overcome this activation energy, one of two methods can be used to facilitate a reaction to take place: increase the temperature of the reaction or add a catalyst.
- **Increasing the temperature of a reaction** means that the average energy of the reactant particles increases and they are more likely to have enough energy to overcome the activation energy.
- A **catalyst** is used to lower the activation energy so that the reaction is more likely to take place. A catalyst does this by providing an alternative, lower energy pathway, for the reaction.
- A catalyst therefore **speeds up a reaction** but does not become part of the reaction in any way.
- **Chemical equilibrium** is the state of a reaction, where the concentrations of the reactants and the products have no net change over time. Usually this occurs when the rate of the forward reaction is the same as the rate of the reverse reaction.
- The **equilibrium constant** relates to reactions at equilibrium, and can be calculated using the following equation:

$$k_c = \frac{[C]^c[D]^d}{[A]^a[B]^b}$$

  where A and B are reactants, C and D are products and a, b, c, and d are the coefficients of the respective reactants and products.
- A **high k_c value** means that the concentration of products at equilibrium is high and the reaction has a high yield. A **low k_c value** means that the concentration of products at equilibrium is low and the reaction has a low yield.
- **Le Chatelier's Principle** states that if a chemical system at equilibrium experiences a change in concentration, temperature or total pressure the equilibrium will shift in order to minimise that change. For example, if the pressure of a gaseous system at eqilibrium was increased, the equilibrium would shift to favour the reaction that produces the lowest quantity of the gas. If the temperature of the same system was to increase, the equilibrium would shift to favour the endothermic reaction. Similar principles apply for changes in concentration of the reactants or products in a reaction.
- The principles of equilibrium are very important in **industrial applications** such as the Haber process, so that productivity can be maximised.

Exercise: Summary Exercise

1. For each of the following questions, choose the one correct answer from the list provided.
   (a) Consider the following reaction that has reached equilibrium after some time in a sealed 1 dm^3 flask:
   $$PCl_5(g) \Leftrightarrow PCl_3(g) + Cl_2(g);\ \Delta H \text{ is positive}$$
   Which one of the following reaction conditions applied to the system would decrease the rate of the reverse reaction?
   i. increase the pressure
   ii. increase the reaction temperature
   iii. continually remove $Cl_2(g)$ from the flask
   iv. addition of a suitable catalyst
   (IEB Paper 2, 2001)
   (b) The following equilibrium constant expression is given for a particular reaction:
   $$K_c = [H_2O]^4[CO_2]^3/[C_3H_8][O_2]^5$$
   For which one of the following reactions is the above expression of K_c is correct?
   i. $C_3H_8(g) + 5O_2(g) \Leftrightarrow 4H_2O(g) + 3CO_2(g)$
   ii. $4H_2O(g) + 3CO_2(g) \Leftrightarrow C_3H_8(g) + 5O_2(g)$
   iii. $2C_3H_8(g) + 7O_2(g) \Leftrightarrow 6CO(g) + 8H_2O(g)$
   iv. $C_3H_8(g) + 5O_2(g) \Leftrightarrow 4H_2O(l) + 3CO_2(g)$
   (IEB Paper 2, 2001)
2. 10 g of magnesium ribbon reacts with a 0.15 mol.dm^{-3} solution of hydrochloric acid at a temperature of 25^{0}C.
   (a) Write a balanced chemical equation for the reaction.
   (b) State **two** ways of increasing the rate of production of $H_2(g)$.
   (c) A table of the results is given below:

Time elapsed (min)	Vol of $H_2(g)$ (cm^3)
0	0
0.5	17
1.0	25
1.5	30
2.0	33
2.5	35
3.0	35

   i. Plot a graph of volume versus time for these results.
   ii. Explain the shape of the graph during the following two time intervals: t = 0 to t = 2.0 min and then t = 2.5 and t = 3.0 min by referring to the volume of $H_2(g)$ produced.
   (IEB Paper 2, 2001)
3. Cobalt chloride crystals are dissolved in a beaker containing ethanol and then a few drops of water are added. After a period of time, the reaction reaches equilibrium as follows:
   $$CoCl_4^{2-} \text{ (blue) } +6H_2O \Leftrightarrow Co(H_2O)_6^{2+} \text{ (pink) } +4Cl^-$$
   The solution, which is now just blue, is poured into three test tubes. State, in each case, what colour changes will be observed (if any) if the following are added in turn to each test tube:
   (a) 1 cm^3 of distilled water
   (b) A few crystals of sodium chloride

(c) The addition of dilute hydrochloric acid to the third test tube causes the solution to turn pink. Explain why this occurs.

(IEB Paper 2, 2001)

Chapter 17

Electrochemical Reactions - Grade 12

17.1 Introduction

Chapter 15 in Grade 11 discussed *oxidation*, *reduction* and *redox reactions*. **Oxidation** involves a **loss of electrons** and **reduction** involves a **gain of electrons**. A **redox reaction** is a reaction where both oxidation and reduction take place. What is common to all of these processes is that they involve a *transfer of electrons* and a change in the oxidation state of the elements that are involved.

?

Exercise: Oxidation and reduction

1. Define the terms **oxidation** and **reduction**.
2. In each of the following reactions say whether the **iron** in the reactants is oxidised or reduced.
   (a) $Fe \rightarrow Fe^{2+} + 2e^-$
   (b) $Fe^{3+} + e^- \rightarrow Fe^{2+}$
   (c) $Fe_2O_3 \rightarrow Fe$
   (d) $Fe^{2+} \rightarrow Fe^{3+} + e^-$
   (e) $Fe_2O_3 + 2Al \rightarrow Al_2$
3. In each of the following equations, say which elements in the reactants are oxidised and which are reduced.
   (a) $CuO(s) + H_2(g) \rightarrow Cu(s) + H_2O(g)$
   (b) $2NO(g) + 2CO(g) \rightarrow N_2(g) + 2CO_2(g)$
   (c) $Mg(s) + FeSO_4(aq) \rightarrow MgSO_4(aq) + Fe(s)$
   (d) $Zn(s) + 2AgNO_3(aq) \rightarrow 2Ag + Zn(NO_3)_2(aq)$
4. Which one of the substances listed below acts as the oxidising agent in the following reaction?

$$3SO_2 + Cr_2O_7^{2-} + 2H^+ \rightarrow 3SO_4^{2-} + 2Cr^{3+} + H_2O$$

   (a) H^+
   (b) Cr^{3+}
   (c) SO_2
   (d) $Cr_2O_7^{2-}$

In Grade 11, an experiment was carried out to see what happened when zinc granules are added to a solution of copper(II) sulphate. In the experiment, the Cu^{2+} ions from the copper(II) sulphate solution were reduced to copper metal, which was then deposited in a layer on the zinc granules. The zinc atoms were oxidised to form Zn^{2+} ions in the solution. The half reactions are as follows:

$$Cu^{2+}(aq) + 2e^- \rightarrow Cu(s) \text{ (reduction half reaction)}$$
$$Zn(s) \rightarrow Zn^{2+}(aq) + 2e^- \text{ (oxidation half reaction)}$$

The overall redox reaction is:

$$Cu^{2+}(aq) + Zn \rightarrow Cu(s) + Zn^{2+}(aq)$$

There was an increase in the temperature of the reaction when you carried out this experiment. Is it possible that this heat energy could be converted into electrical energy? In other words, can we use a chemical reaction where there is an exchange of electrons, to produce electricity? And if this is possible, what would happen if an electrical current was supplied to *cause* some type of chemical reaction to take place?

An **electrochemical reaction** is a chemical reaction that *produces a voltage*, and therefore a flow of electrical current. An electrochemical reaction can also be the reverse of this process, in other words if an electrical current causes a chemical reaction to take place.

Definition: Electrochemical reaction
If a chemical reaction is caused by an external voltage, or if a voltage is caused by a chemical reaction, it is an electrochemical reaction.

Electrochemistry is the branch of chemistry that studies these electrochemical reactions. In this chapter, we will be looking more closely at different types of electrochemical reactions, and how these can be used in different ways.

17.2 The Galvanic Cell

Activity :: Experiment : Electrochemical reactions
Aim:

To investigate the reactions that take place in a zinc-copper cell
Apparatus:

zinc plate, copper plate, measuring balance, zinc sulphate ($ZnSO_4$) solution (1 mol.dm^{-3}), copper sulphate ($CuSO_4$) solution (1 mol.dm^{-3}), two 250 ml beakers, U-tube, Na_2SO_4 solution, cotton wool, ammeter, connecting wire.
Method:

1. Measure the mass of the copper and zinc plates and record your findings.
2. Pour about 200 ml of the zinc sulphate solution into a beaker and put the zinc plate into it.
3. Pour about 200 ml of the copper sulphate solution into the second beaker and place the copper plate into it.
4. Fill the U-tube with the Na_2SO_4 solution and seal the ends of the tubes with the cotton wool. This will stop the solution from flowing out when the U-tube is turned upside down.

5. Connect the zinc and copper plates to the ammeter and observe whether the ammeter records a reading.
6. Place the U-tube so that one end is in the copper sulphate solution and the other end is in the zinc sulphate solution. Is there a reading on the ammeter? In which direction is the current flowing?
7. Take the ammeter away and connect the copper and zinc plates to each other directly using copper wire. Leave to stand for about one day.
8. After a day, remove the two plates and rinse them first with distilled water, then with alcohol and finally with ether. Dry the plates using a hair dryer.
9. Weigh the zinc and copper plates and record their mass. Has the mass of the plates changed from the original measurements?

Note: A voltmeter can also be used in place of the ammeter. A voltmeter will measure the potential difference across the cell.

Results:

During the experiment, you should have noticed the following:

- When the U-tube containing the Na_2SO_4 solution was absent, there was no reading on the ammeter.
- When the U-tube was connected, a reading was recorded on the ammeter.
- After the plates had been connected directly to each other and left for a day, there was a change in their mass. The mass of the zinc plate decreased, while the mass of the copper plate increased.
- The direction of electron flow is from the zinc plate towards the copper plate.

Conclusions:

When a zinc sulphate solution containing a zinc plate is connected by a U-tube to a copper sulphate solution containing a copper plate, reactions occur in both solutions. The decrease in mass of the zinc plate suggests that the zinc metal has been oxidised. The increase in mass of the copper plate suggests that reduction has occurred here to produce more copper metal. This will be explained in detail below.

17.2.1 Half-cell reactions in the Zn-Cu cell

The experiment above demonstrated a zinc-copper cell. This was made up of a zinc **half cell** and a copper **half cell**.

Definition: Half cell
A half cell is a structure that consists of a conductive electrode surrounded by a conductive electrolyte. For example, a zinc half cell could consist of a zinc metal plate (the electrode) in a zinc sulphate solution (the electrolyte).

How do we explain what has just been observed in the zinc-copper cell?

- *Copper plate*

  At the copper plate, there was an *increase* in mass. This means that Cu^{2+} ions from the copper sulphate solution were deposited onto the plate as atoms of copper metal. The half-reaction that takes place at the copper plate is:

  $$Cu^{2+} + 2e^- \rightarrow Cu \textbf{ (Reduction half reaction)}$$

  Another shortened way to represent this copper half-cell is Cu^{2+}/Cu.

- *Zinc plate*

  At the zinc plate, there was a *decrease* in mass. This means that some of the zinc goes into solution as Z^{2+} ions. The electrons remain on the zinc plate, giving it a negative charge. The half-reaction that takes place at the zinc plate is:

  $$Zn \rightarrow Zn^{2+} + 2e^- \textbf{ (Oxidation half reaction)}$$

  The shortened way to represent the zinc half-cell is Zn/Zn^{2+}.

  The overall reaction is:

  $$Zn + Cu^{2+} + 2e^- \rightarrow Zn^{2+} + Cu + 2e^- \text{ or, if we cancel the electrons:}$$
  $$Zn + Cu^{2+} \rightarrow Zn^{2+} + Cu$$

  For this electrochemical cell, the standard notation is:

  $$Zn|Zn^{2+}||Cu^{2+}|Cu$$

  where

  $$\begin{aligned} | &= \text{a phase boundary (solid/aqueous)} \\ || &= \text{the salt bridge} \end{aligned}$$

In the notation used above, the oxidation half-reaction at the anode is written on the left, and the reduction half-reaction at the cathode is written on the right. In the Zn-Cu electrochemical cell, the direction of current flow in the external circuit is from the zinc electrode (where there has been a build up of electrons) to the copper electrode.

17.2.2 Components of the Zn-Cu cell

In the zinc-copper cell, the copper and zinc plates are called the **electrodes**. The electrode where oxidation occurs is called the **anode**, and the electrode where reduction takes place is called the **cathode**. In the zinc-copper cell, the zinc plate is the anode and the copper plate is the cathode.

Definition: Electrode
An electrode is an electrical conductor that is used to make contact with a metallic part of a circuit. The anode is the electrode where oxidation takes place. The cathode is the electrode where reduction takes place.

The zinc sulphate and copper sulphate solutions are called the **electrolyte** solutions.

Definition: Electrolyte
An electrolyte is a substance that contains free ions and which therefore behaves as an electrical conductor.

The U-tube also plays a very important role in the cell. In the Zn/Zn^{2+} half-cell, there is a build up of positive charge because of the release of electrons through oxidation. In the Cu^{2+}/Cu half-cell, there is a decrease in the positive charge because electrons are gained through reduction. This causes a movement of SO_4^{2-} ions into the beaker where there are too many positive ions, in order to neutralise the solution. Without this, the flow of electrons in the outer circuit stops completely. The U-tube is called the **salt bridge**. The salt bridge acts as a transfer medium that allows ions to flow through without allowing the different solutions to mix and react.

Definition: Salt bridge
A salt bridge, in electrochemistry, is a laboratory device that is used to connect the oxidation and reduction half-cells of a galvanic cell.

17.2.3 The Galvanic cell

In the zinc-copper cell the important thing to notice is that the chemical reactions that take place at the two electrodes cause an electric current to flow through the outer circuit. In this type of cell, **chemical energy is converted to electrical energy**. These are called **galvanic cells**. The zinc-copper cell is one example of a galvanic cell. A galvanic cell (which is also sometimes referred to as a **voltaic** or **electrochemical** cell) consists of two metals that are connected by a salt bridge between the individual half-cells. A galvanic cell generates electricity using the reactions that take place at these two metals, each of which has a different **reaction potential**.

So what is meant by the 'reaction potential' of a substance? Every metal has a different half reaction and different dissolving rates. When two metals with different reaction potentials are used in a galvanic cell, a potential difference is set up between the two electrodes, and the result is a flow of current through the wire that connects the electrodes. In the zinc-copper cell, zinc has a higher reaction potential than copper and therefore dissolves more readily into solution. The metal 'dissolves' when it loses electrons to form positive metal ions. These electrons are then transferred through the connecting wire in the outer circuit.

Definition: Galvanic cell
A galvanic (voltaic) cell is an electrochemical cell that uses a chemical reaction between two dissimilar electrodes dipped in an electrolyte, to generate an electric current.

Interesting Fact

It was the Italian physician and anatomist Luigi Galvani who marked the birth of electrochemistry by making a link between chemical reactions and electricity. In 1780, Galvani discovered that when two different metals (copper and zinc for example) were connected together and then both touched to different parts of a nerve of a frog leg at the same time, they made the leg contract. He called this "animal electricity". While many scientists accepted his ideas, another scientist, Alessandro Volta, did not. In 1800, because of his professional disagreement over the galvanic response that had been suggested by Luigi Galvani, Volta developed the *voltaic pile*, which was very similar to the galvanic cell. It was the work of these two men that paved the way for all electrical batteries.

Worked Example 83: Understanding galvanic cells

Question: For the following cell:

$$Zn|Zn^{2+}||Ag^{+}|Ag$$

1. Give the anode and cathode half-reactions.
2. Write the overall equation for the chemical reaction.
3. Give the direction of the current in the external circuit.

Answer

Step 1 : Identify the oxidation and reduction reactions

In the standard notation format, the oxidation reaction is written on the left and the reduction reaction on the right. So, in this cell, zinc is oxidised and silver ions are reduced.

Step 2 : Write the two half reactions

Oxidation half-reaction:
$Zn \rightarrow Zn^{2+} + 2e^{-}$

Reduction half-reaction:
$Ag^{+} + e^{-} \rightarrow Ag$

Step 3 : Combine the half-reactions to get the overall equation.

When you combine the two half-reactions, all the reactants must go on the left side of the equation and the products must go on the right side of the equation. The overall equation therefore becomes:

$$\mathrm{Zn + Ag^{+} + e^{-} \rightarrow Zn^{2+} + 2e^{-} + Ag}$$

Note that this equation is not balanced. This will be discussed later in the chapter.

Step 4 : Determine the direction of current flow

A build up of electrons occurs where oxidation takes place. This is at the zinc electrode. Current will therefore flow from the zinc electrode to the silver electrode.

17.2.4 Uses and applications of the galvanic cell

The principles of the galvanic cell are used to make **electrical batteries**. In science and technology, a battery is a device that stores chemical energy and makes it available in an electrical form. Batteries are made of electrochemical devices such as one or more galvanic cells, fuel cells or flow cells. Batteries have many uses including in torches, electrical appliances (long-life alkaline batteries), digital cameras (lithium battery), hearing aids (silver-oxide battery), digital watches (mercury battery) and military applications (thermal battery). Refer to chapter 23 for more information on batteries.

The galvanic cell can also be used for **electroplating**. Electroplating occurs when an electrically conductive object is coated with a layer of metal using electrical current. Sometimes, electroplating is used to give a metal particular properties such as corrosion protection or wear resistance. At other times, it can be for aesthetic reasons for example in the production of jewellery. This will be discussed in more detail later in this chapter.

Exercise: Galvanic cells

1. The following half-reactions take place in an electrochemical cell:
   $\mathrm{Fe} \rightarrow \mathrm{Fe}^{3+} + 3\mathrm{e}^-$
   $\mathrm{Fe}^{2+} + 2\mathrm{e}^- \rightarrow \mathrm{Fe}$
   (a) Which is the oxidation half-reaction?
   (b) Which is the reduction half-reaction?
   (c) Name one oxidising agent.
   (d) Name one reducing agent.
   (e) Use standard notation to represent this electrochemical cell.
2. For the following cell:
   $$Mg|Mg^{2+}||Mn^{2+}|Mn$$
   (a) Give the cathode half-reaction.
   (b) Give the anode half-reaction.
   (c) Give the overall equation for the electrochemical cell.
   (d) What metals could be used for the electrodes in this electrochemical cell.
   (e) Suggest two electrolytes for this electrochemical cell.
   (f) In which direction will the current flow?
   (g) Draw a simple sketch of the complete cell.
3. For the following cell:
   $$Sn|Sn^{2+}||Ag^{+}|Ag$$
   (a) Give the cathode half-reaction.
   (b) Give the anode half-reaction.
   (c) Give the overall equation for the electrochemical cell.
   (d) Draw a simple sketch of the complete cell.

17.3 The Electrolytic cell

In section 17.2, we saw that a chemical reaction that involves a transfer of electrons, can be used to produce an electric current. In this section, we are going to see whether the 'reverse' process applies. In other words, is it possible to use an electric current to force a particular chemical reaction to occur, which would otherwise not take place? The answer is 'yes', and the type of cell that is used to do this, is called an **electrolytic cell**.

Definition: Electrolytic cell
An electrolytic cell is a type of cell that uses electricity to drive a non-spontaneous reaction.

An electrolytic cell is activated by applying an electrical potential across the anode and cathode to force an internal chemical reaction between the ions that are in the electrolyte solution. This process is called **electrolysis**.

Definition: Electrolysis
In chemistry and manufacturing, electrolysis is a method of separating bonded elements and compounds by passing an electric current through them.

Activity :: Demonstration : The movement of coloured ions

A piece of filter paper is soaked in an ammonia-ammonium chloride solution and placed on a microscope slide. The filter paper is then connected to a supply of electric current using crocodile clips and connecting wire as shown in the diagram below. A line of copper chromate solution is placed in the centre of the filter paper. The colour of this solution is initially green-brown.

The current is then switched on and allowed to run for about 20 minutes. After this time, the central coloured band disappears and is replaced by two bands, one yellow and the other blue, which seem to have separated out from the first band of copper chromate.

Explanation:

- The cell that is used to supply an electric current sets up a potential difference across the circuit, so that one of the electrodes is positive and the other is negative.
- The chromate (CrO_4^{2-}) ions in the copper chromate solution are attracted to the positive electrode, while the Cu^{2+} ions are attracted to the negative electrode.

Conclusion:

The movement of ions occurs because the electric current in the outer circuit sets up a potential difference between the two electrodes.

Similar principles apply in the electrolytic cell, where substances that are made of ions can be broken down into simpler substances through electrolysis.

17.3.1 The electrolysis of copper sulphate

There are a number of examples of electrolysis. The electrolysis of copper sulphate is just one.

Activity :: Demonstration : The electrolysis of copper sulphate

Two copper electrodes are placed in a solution of blue copper sulphate and are connected to a source of electrical current as shown in the diagram below. The current is turned on and the reaction is left for a period of time.

Observations:

- The initial blue colour of the solution remains unchanged.
- It appears that copper has been *deposited* on one of the electrodes but *dissolved* from the other.

Explanation:

- At the negative cathode, positively charged Cu^{2+} ions are attracted to the negatively charged electrode. These ions gain electrons and are reduced to form copper metal, which is deposited on the electrode. The half-reaction that takes place is as follows:

$$Cu^{2+}(aq) + 2e^- \rightarrow Cu(s) \textbf{ (reduction half reaction)}$$

- At the positive anode, copper metal is oxidised to form Cu^{2+} ions. This is why it appears that some of the copper has dissolved from the electrode. The half-reaction that takes place is as follows:

$$Cu(s) \rightarrow Cu^{2+}(aq) + 2e^- \textbf{ (oxidation half reaction)}$$

- The amount of copper that is *deposited* at one electrode is approximately the same as the amount of copper that is *dissolved* from the other. The number of Cu^{2+} ions in the solution therefore remains almost the same and the blue colour of the solution is unchanged.

Conclusion:

In this demonstration, an electric current was used to split $CuSO_4$ into its component ions, Cu^{2+} and SO_4^{2-}. This process is called *electrolysis*.

17.3.2 The electrolysis of water

Water can also undergo electrolysis to form hydrogen gas and oxygen gas according to the following reaction:

$$2H_2O(l) \rightarrow 2H_2(g) + O_2(g)$$

This reaction is very important because hydrogen gas has the potential to be used as an energy source. The electrolytic cell for this reaction consists of two electrodes (normally platinum metal), submerged in an electrolyte and connected to a source of electric current.

The reduction half-reaction that takes place at the cathode is as follows:

$$2H_2O(l) + 2e^- \rightarrow H_2(g) + 2OH^-(aq)$$

The oxidation half-reaction that takes place at the anode is as follows:

$$2H_2O(l) \rightarrow O_2(g) + 4H^+(aq) + 4e^-$$

17.3.3 A comparison of galvanic and electrolytic cells

It should be much clearer now that there are a number of differences between a *galvanic* and an *electrolytic* cell. Some of these differences have been summarised in table 17.1.

Item	**Galvanic cell**	**Electrolytic cell**
Metals used for electrode	Two metals with different reaction potentials are used as electrodes	The same metal can be used for both the cathode and the anode
Charge of the anode	negative	positive
Charge of the cathode	positive	negative
The electrolyte solution/s	The electrolyte solutions are kept separate from one another, and are connected only by a salt bridge	The cathode and anode are in the same electrolyte
Energy changes	Chemical potential energy from chemical reactions is converted to electrical energy	An external supply of electrical energy causes a chemical reaction to occur
Applications	Run batteries, electroplating	Electrolysis e.g. of water, NaCl

Table 17.1: A comparison of galvanic and electrolytic cells

?

Exercise: Electrolyis

1. An electrolytic cell consists of two electrodes in a silver chloride (AgCl) solution, connected to a source of current. A current is passed through the solution and Ag^+ ions are reduced to a silver metal deposit on one of the electrodes.
   (a) Give the equation for the reduction half-reaction.
   (b) Give the equation for the oxidation half-reacion.
2. Electrolysis takes place in a solution of molten lead bromide (PbBr) to produce lead atoms.
   (a) Draw a simple diagram of the electrolytic cell.
   (b) Give equations for the half-reactions that take place at the anode and cathode, and include these in the diagram.
   (c) On your diagram, show the direction in which current flows.

17.4 Standard Electrode Potentials

If a voltmeter is connected in the circuit of an electrochemical cell, a reading is obtained. In other words, there is a **potential difference** between the two half cells. In this section, we are going to look at this in more detail to try to understand more about the **electrode potentials** of each of the electrodes in the cell. We are going to break this section down so that you build up your understanding gradually. Make sure that you understand each subsection fully before moving on, otherwise it might get confusing!

17.4.1 The different reactivities of metals

All metals have different reactivities. When metals react, they give away electrons and form positive ions. But some metals do this more easily than others. Look at the following two half reactions:

$$Zn \rightarrow Zn^{2+} + 2e^-$$
$$Cu \rightarrow Cu^{2+} + 2e^-$$

Of these two metals, zinc is more reactive and is more likely to give away electrons to form Zn^{2+} ions in solution, than is copper.

17.4.2 Equilibrium reactions in half cells

Let's think back to the Zn-Cu electrochemical cell. This cell is made up of two half cells and the reactions that take place at each of the electrodes are as follows:

$$Zn \rightarrow Zn^{2+} + 2e^-$$
$$Cu^{2+} + 2e^- \rightarrow Cu$$

At the zinc electrode, the zinc metal loses electrons and forms Zn^{2+} ions. The electrons are concentrated on the zinc metal while the Zn^{2+} ions are in solution. But some of the ions will be attracted back to the negatively charged metal, will gain their electrons again and will form zinc metal. A **dynamic equilibrium** is set up between the zinc metal and the Zn^{2+} ions in solution when the rate at which ions are *leaving* the metal is equal to the rate at which they are *joining* it again. The situation looks something like the diagram in figure 17.1.

Figure 17.1: Zinc loses electrons to form positive ions in solution. The electrons accumulate on the metal surface.

The equilibrium reaction is represented like this:

$$Zn^{2+}(aq) + 2e^- \Leftrightarrow Zn(s)$$

(NOTE: By convention, the ions are written on the *left* hand side of the equation)

In the zinc half cell, the equilibrium lies far to the left because the zinc loses electrons easily to form Zn^{2+} ions. We can also say that the zinc is *oxidised* and that it is a strong *reducing agent*.

At the copper electrode, a similar process takes place. The difference though is that copper is not as reactive as zinc and so it does not form ions as easily. This means that the build up of electrons on the copper electrode is less (figure 17.2).

The equilibrium reaction is shown like this:

Figure 17.2: Zinc loses electrons to form positive ions in solution. The electrons accumulate on the metal surface.

$$Cu^{2+}(aq) + 2e^- \Leftrightarrow Cu(s)$$

The equation lies far to the right because most of the copper is present as copper metal rather than as Cu^{2+} ions. In this half reaction, the Cu^{2+} ions are *reduced*.

17.4.3 Measuring electrode potential

If we put the two half cells together, a potential difference is set up in two places in the Zn-Cu cell:

1. There is a potential difference between the metal and the solution surrounding it because one is more negative than the other.
2. There is a potential difference between the Zn and Cu electrodes because one is more negative than the other.

It is the **potential difference** (recorded as a voltage) between the two electrodes that causes electrons, and therefore current, to flow from the more negative electrode to the less negative electrode.

The problem though is that we cannot measure the potential difference (voltage) between a metal and its surrounding solution in the cell. To do this, we would need to connect a voltmeter to both the metal and the solution, which is not possible. This means we cannot measure the exact **electrode potential** (E^oV) of a particular metal. The electrode potential describes the ability of a metal to give up electrons. And if the exact electrode potential of each of the electrodes involved can't be measured, then it is difficult to calculate the potential difference between them. But what we *can* do is to try to describe the electrode potential of a metal *relative* to another substance. We need to use a **standard reference electrode** for this.

17.4.4 The standard hydrogen electrode

Before we look at the standard hydrogen electrode, it may be useful to have some more understanding of the ideas behind a 'reference electrode'. Refer to the Tip box on 'Understanding the ideas behind a reference electrode' before you read further.

Important: Understanding the ideas behind a reference electrode

Adapted from www.chemguide.co.uk

Let's say that you have a device that you can use to measure heights from some distance away. You want to use this to find out how tall a particular person is. Unfortunately, you can't see their feet because they are standing in long grass. Although you can't measure their absolute height, what you can do is to measure their height relative to the post next to them. Let's say that person A for example is 15 cm *shorter* than the height of the post. You could repeat this for a number of other people (B and C). Person B is 30 cm shorter than the post and person C is 10 cm *taller* than the post.

You could summarise your findings as follows:

Person	Height relative to post (cm)
A	-15
B	-30
C	+10

Although you don't know any of their absolute heights, you can rank them in order, and do some very simple sums to work out exactly how much taller one is than another. For example, person C is 25 cm taller than A and 40 cm taller than B.

As mentioned earlier, it is difficult to measure the *absolute* electrode potential of a particular substance, but we can use a reference electrode (similar to the 'post' in the Tip box example) that we use to calculate *relative* electrode potentials for these substances. The reference elctrode that is used is the **standard hydrogen electrode** (figure 17.3).

Definition: Standard hydrogen electrode

The standard hydrogen electrode is a redox electrode which forms the basis of the scale of oxidation-reduction potentials. The actual electrode potential of the hydrogen electrode is estimated to be 4.44 0.02 V at 25^0C, but its standard electrode potential is said to be zero at all temperatures so that it can be used as for comparison with other electrodes. The hydrogen electrode is based on the following redox half cell:

$$2H^+(aq) + 2e^- \rightarrow H_2(g)$$

A standard hydrogen electrode consists of a platinum electrode in a solution containing H^+ ions. The solution (e.g. H_2SO_4) that contains the H^+ ions has a concentration of 1 mol.dm^{-3}. As the hydrogen gas bubbles over the platinum electrode, an equilibrium is set up between hydrogen molecules and hydrogen ions in solution. The reaction is as follows:

Figure 17.3: The standard hydrogen electrode

$$2H^+(aq) + 2e^- \Leftrightarrow H_2(g)$$

The position of this equilibrium can change if you change some of the conditions (e.g. concentration, temperature). It is therefore important that the conditions for the standard hydrogen electrode are standardised as follows: pressure = 100 kPa (1atm); temperature = 298 K (25^0C) and concentration = 1 mol.dm^{-3}.

In order to *use* the hydrogen electrode, it needs to be attached to the electrode system that you are investigating. For example, if you are trying to determine the electrode potential of copper, you will need to connect the copper half cell to the hydrogen electrode; if you are trying to determine the electrode potential of zinc, you will need to connect the zinc half cell to the hydrogen electrode and so on. Let's look at the examples of zinc and copper in more detail.

1. **Zinc**

   Zinc has a greater tendency than hydrogen to form ions, so if the standard hydrogen electrode is connected to the zinc half cell, the zinc will be relatively more negative because the electrons that are released when zinc is oxidised will accumulate on the metal. The equilibria on the two electrodes are as follows:

   $$Zn^{2+}(aq) + 2e^- \Leftrightarrow Zn(s)$$
   $$2H^+(aq) + 2e^- \Leftrightarrow H_2(g)$$

   In the zinc half-reaction, the equilibrium lies far to the left and in the hydrogen half-reaction, the equilibrium lies far to the right. A simplified representation of the cell is shown in figure 17.4.

   The voltmeter measures the potential difference between the charge on these electrodes. In this case, the voltmeter would read 0.76 and would show that Zn is the negative electrode (i.e. it has a relatively higher number of electrons).

2. **Copper**

   Copper has a lower tendency than hydrogen to form ions, so if the standard hydrogen electrode is connected to the copper half cell, the hydrogen will be relatively more negative. The equilibria on the two electrodes are as follows:

Figure 17.4: When zinc is connected to the standard hydrogen electrode, relatively few electrons build up on the platinum (hydrogen) electrode. There are lots of electrons on the zinc electrode.

$$Cu^{2+}(aq) + 2e^- \Leftrightarrow Cu(s)$$
$$2H^+(aq) + 2e^- \Leftrightarrow H_2(g)$$

In the copper half-reaction, the equilibrium lies far to the right and in the hydrogen half-reaction, the equilibrium lies far to the left. A simplified representation of the cell is shown in figure 17.5.

Figure 17.5: When copper is connected to the standard hydrogen electrode, relatively few electrons build up on the copper electrode. There are lots of electrons on the hydrogen electrode.

The voltmeter measures the potential difference between the charge on these electrodes. In this case, the voltmeter would read 0.34 and would show that Cu is the positive electrode (i.e. it has a relatively lower number of electrons).

17.4.5 Standard electrode potentials

The voltages recorded earlier when zinc and copper were connected to a standard hydrogen electrode are in fact the **standard electrode potentials** for these two metals. It is important to remember that these are not *absolute* values, but are potentials that have been measured *relative* to the potential of hydrogen if the standard hydrogen electrode is taken to be zero.

Important: Conventions and voltage sign

By convention, the hydrogen electrode is written on the *left hand side* of the cell. The sign of the voltage tells you the sign of the metal electrode.

In the examples we used earlier, zinc's electrode potential is actually -0.76 and copper is +0.34. So, if a metal has a *negative* standard electrode potential, it means it forms ions easily. The more negative the value, the easier it is for that metal to form ions. If a metal has a *positive* standard electrode potential, it means it does not form ions easily. This will be explained in more detail below.

Luckily for us, we do not have to calculate the standard electrode potential for every metal. This has been done already and the results are recorded in a table of standard electrode potentials (table 17.2).

A few examples from the table are shown in table 17.3. These will be used to explain some of the trends in the table of electrode potentials.

Refer to table 17.3 and notice the following trends:

- Metals at the top of series (e.g. Li) have more negative values. This means they ionise easily, in other words, they release electrons easily. These metals are easily **oxidised** and are therefore good **reducing agents**.
- Metal ions at the bottom of the table are good at picking up electrons. They are easily **reduced** and are therefore good **oxidising agents**.
- The reducing ability (i.e. the ability to act as a reducing agent) of the metals in the table *increases* as you move *up* in the table.
- The oxidising ability of metals *increases* as you move *down* in the table.

Worked Example 84: Using the table of Standard Electrode Potentials

Question:
The following half-reactions take place in an electrochemical cell:
$Cu^{2+} + 2e^- \Leftrightarrow Cu$
$Ag^- + e^- \Leftrightarrow Ag$

1. Which of these reactions will be the oxidation half-reaction in the cell?
2. Which of these reactions will be the reduction half-reaction in the cell?

Answer
Step 5 : Determine the electrode potential for each metal
From the table of standard electrode potentials, the electrode potential for the copper half-reaction is +0.34 V. The electrode potential for the silver half-reaction is +0.80 V.

Step 6 : Use the electrode potential values to determine which metal is oxidised and which is reduced
Both values are positive, but silver has a higher positive electrode potential than copper. This means that silver does not form ions easily, in other words, silver is more likely to be *reduced*. Copper is more likely to be *oxidised* and to form ions more easily than silver. Copper is the oxidation half-reaction and silver is the reduction half-reaction.

Half-Reaction	E^0V
$Li^+ + e^- \rightleftharpoons Li$	-3.04
$K^+ + e^- \rightleftharpoons K$	-2.92
$Ba^{2+} + 2e^- \rightleftharpoons Ba$	-2.90
$Ca^{2+} + 2e^- \rightleftharpoons Ca$	-2.87
$Na^+ + e^- \rightleftharpoons Na$	-2.71
$Mg^{2+} + 2e^- \rightleftharpoons Mg$	-2.37
$Mn^{2+} + 2e^- \rightleftharpoons Mn$	-1.18
$2H2O + 2e^- \rightleftharpoons H_2(g) + 2OH^-$	-0.83
$Zn^{2+} + 2e^- \rightleftharpoons Zn$	-0.76
$Cr^{2+} + 2e^- \rightleftharpoons Cr$	-0.74
$Fe^{2+} + 2e^- \rightleftharpoons Fe$	-0.44
$Cr^{3+} + 3e^- \rightleftharpoons Cr$	-0.41
$Cd^{2+} + 2e^- \rightleftharpoons Cd$	-0.40
$Co^{2+} + 2e^- \rightleftharpoons Co$	-0.28
$Ni^{2+} + 2e^- \rightleftharpoons Ni$	-0.25
$Sn^{2+} + 2e^- \rightleftharpoons Sn$	-0.14
$Pb^{2+} + 2e^- \rightleftharpoons Pb$	-0.13
$Fe^{3+} + 3e^- \rightleftharpoons Fe$	-0.04
$2H^+ + 2e^- \rightleftharpoons H_2(g)$	**0.00**
$S + 2H^+ + 2e^- \rightleftharpoons H_2S(g)$	0.14
$Sn^{4+} + 2e^- \rightleftharpoons Sn^{2+}$	0.15
$Cu^{2+} + e^- \rightleftharpoons Cu^+$	0.16
$SO_4^{2+} + 4H^+ + 2e^- \rightleftharpoons SO_2(g) + 2H_2O$	0.17
$Cu^{2+} + 2e^- \rightleftharpoons Cu$	0.34
$2H_2O + O_2 + 4e^- \rightleftharpoons 4OH^-$	0.40
$Cu^+ + e^- \rightleftharpoons Cu$	0.52
$I_2 + 2e^- \rightleftharpoons 2I^-$	0.54
$O_2(g) + 2H^+ + 2e^- \rightleftharpoons H_2O_2$	0.68
$Fe^{3+} + e^- \rightleftharpoons Fe^{2+}$	0.77
$NO_3^- + 2H^+ + e^- \rightleftharpoons NO_2(g) + H_2O$	0.78
$Hg^{2+} + 2e^- \rightleftharpoons Hg(l)$	0.78
$Ag^+ + e^- \rightleftharpoons Ag$	0.80
$NO_3^- + 4H^+ + 3e^- \rightleftharpoons NO(g) + 2H_2O$	0.96
$Br_2 + 2e^- \rightleftharpoons 2Br^-$	1.06
$O_2(g) + 4H^+ + 4e^- \rightleftharpoons 2H_2O$	1.23
$MnO_2 + 4H^+ + 2e^- \rightleftharpoons Mn^{2+} + 2H_2O$	1.28
$Cr_2O_7^{2-} + 14H^+ + 6e^- \rightleftharpoons 2Cr^{3+} + 7H_2O$	1.33
$Cl_2 + 2e^- \rightleftharpoons 2Cl^-$	1.36
$Au^{3+} + 3e^- \rightleftharpoons Au$	1.50
$MnO_4^- + 8H^+ + 5e^- \rightleftharpoons Mn^{2+} + 4H_2O$	1.52
$Co^{3+} + e^- \rightleftharpoons Co^{2+}$	1.82
$F_2 + 2e^- \rightleftharpoons 2F^-$	2.87

Table 17.2: Standard Electrode Potentials

Half-Reaction	E^0V
$Li^+ + e^- \rightleftharpoons Li$	-3.04
$Zn^{2+} + 2e^- \rightleftharpoons Zn$	-0.76
$Fe^{3+} + 3e^- \rightleftharpoons Fe$	-0.04
$2H^+ + 2e^- \rightleftharpoons H_2(g)$	**0.00**
$Cu^{2+} + 2e^- \rightleftharpoons Cu$	0.34
$Hg^{2+} + 2e^- \rightleftharpoons Hg(l)$	0.78
$Ag^+ + e^- \rightleftharpoons Ag$	0.80

Table 17.3: A few examples from the table of standard electrode potentials

Important: Learning to understand the question in a problem.
Before you tackle this problem, make sure you understand exactly what the question is asking. If magnesium is able to displace silver from a solution of silver nitrate, this means that magnesium metal will form magnesium ions and the silver ions will become silver metal. In other words, there will now be *silver* metal and a solution of *magnesium nitrate*. This will only happen if magnesium has a greater tendency than silver to form ions. In other words, what the question is actually asking is whether magnesium or silver forms ions more easily.

Worked Example 85: Using the table of Standard Electrode Potentials

Question: Is magnesium able to displace silver from a solution of silver nitrate?
Answer
Step 1 : Determine the half-reactions that would take place if magnesium were to displace silver nitrate.
The half-reactions are as follows:
$Mg^{2+} + 2e^- \Leftrightarrow Mg$
$Ag^+ + e^- \Leftrightarrow Ag$

Step 2 : Use the table of electrode potentials to see which metal forms ions more easily.
Looking at the electrode potentials for the magnesium and silver reactions:
For the magnesium half-reaction: E^{o}V = -2.37
For the silver half-reaction: E^{o}V = 0.80

This means that magnesium is more easily **oxidised** than silver and the equilibrium in this half-reaction lies to the left. The oxidation reaction will occur spontaneously in magnesium. Silver is more easily **reduced** and the equilibrium lies to the right in this half-reaction. It can be concluded that magnesium will displace silver from a silver nitrate solution so that there is silver metal and magnesium ions in the solution.

Exercise: Table of Standard Electrode Potentials

1. In your own words, explain what is meant by the 'electrode potential' of a metal.
2. Give the standard electrode potential for each of the following metals:
   (a) magnesium

(b) lead
(c) nickel

3. Refer to the electrode potentials in table 17.3.
   (a) Which of the metals is most likely to be oxidised?
   (b) Which metal is most likely to be reduced?
   (c) Which metal is the strongest reducing agent?
   (d) In the copper half-reaction, does the equilibrium position for the reaction lie to the left or to the right? Explain your answer.
   (e) In the mercury half-reaction, does the equilibrium position for the reaction lie to the left or to the right? Explain your answer.
   (f) If silver was added to a solution of copper sulphate, would it displace the copper from the copper sulphate solution? Explain your answer.
4. Use the table of standard electrode potentials to put the following in order from the *strongest oxidising agent* to the *weakest oxidising agent*.
   - Cu^{2+}
   - MnO_4^-
   - Br_2
   - Zn^{2+}
5. Look at the following half-reactions.
   - $Ca^{2+} + 2e^- \rightarrow Ca$
   - $Cl_2 + 2e^- \rightarrow 2Cl$
   - $Fe^{3+} + 3e^- \rightarrow Fe$
   - $I_2 + 2e^- \rightarrow 2I^-$

   (a) Which substance is the strongest oxidising agent?
   (b) Which substance is the strongest reducing agent?
6. Which one of the substances listed below acts as the oxidising agent in the following reaction?

$$3SO_2 + Cr_2O_7^{2-} + 2H^+ \rightarrow 3SO_4^{2-} + 2Cr^{3+} + H_2O$$

   (a) H^+
   (b) Cr^{3+}
   (c) SO_2
   (d) $Cr_2O_7^{2-}$

   (IEB Paper 2, 2004)
7. If zinc is added to a solution of magnesium sulphate, will the zinc displace the magnesium from the solution? Give a detailed explanation for your answer.

17.4.6 Combining half cells

Let's stay with the example of the zinc and copper half cells. If we combine these cells as we did earlier in the chapter (section 17.2), the following two equilibria exist:

$$Zn^{2+} + 2e^- \Leftrightarrow Zn(E^0 = -0.76V)$$
$$Cu^{2+} + 2e^- \Leftrightarrow Cu(E^0 = +0.34V)$$

We know from demonstrations, and also by looking at the sign of the electrode potential, that when these two half cells are combined, zinc will be the oxidation half-reaction and copper will be the reduction half-reaction. A voltmeter connected to this cell will show that the zinc electrode is more negative than the copper electrode. The reading on the meter will show the potential difference between the two half cells. This is known as the **electromotive force (emf)** of the cell.

Definition: Electromotive Force (emf)
The emf of a cell is defined as the maximum potential difference between two electrodes or half cells in a voltaic cell. emf is the electrical driving force of the cell reaction. In other words, the higher the emf, the stronger the reaction.

Definition: Standard emf (E^0_{cell})
Standard emf is the emf of a voltaic cell operating under standard conditions (i.e. 100 kPa, concentration = 1 mol.dm^{-3} and temperature = 298 K). The symbol ^{0} denotes standard conditions.

When we want to represent this cell, it is shown as follows:

$$Zn|Zn^{2+}(1mol.dm^{-3})||Cu^{2+}(1mol.dm^{-3})|Cu$$

The **anode** half cell (where *oxidation* takes place) is always written on the *left*. The **cathode** half cell (where *reduction* takes place) is always written on the *right*.

It is important to note that the potential difference across a cell is related to the extent to which the spontaneous cell reaction has reached equilibrium. In other words, as the reaction proceeds and the concentration of reactants decreases and the concentration of products increases, the reaction approaches equilibrium. When equilibrium is reached, the emf of the cell is zero and the cell is said to be 'flat'. There is no longer a potential difference between the two half cells, and therefore no more current will flow.

17.4.7 Uses of standard electrode potential

Standard electrode potentials have a number of different uses.

Calculating the emf of an electrochemical cell

To calculate the emf of a cell, you can use any one of the following equations:

$E^0_{(cell)}$ = E^0 (right) - E^0 (left) ('right' refers to the electrode that is written on the right in standard cell notation. 'Left' refers to the half-reaction written on the left in this notation)

$E^0_{(cell)}$ = E^0 (reduction half reaction) - E^0 (oxidation half reaction)

$E^0_{(cell)}$ = E^0 (oxidising agent) - E^0 (reducing agent)

$E^0_{(cell)}$ = E^0 (cathode) - E^0 (anode)

So, for the Zn-Cu cell,

$E^0_{(cell)}$ = 0.34 - (-0.76)

= 0.34 + 0.76

= 1.1 V

Worked Example 86: Calculating the emf of a cell

Question: The following reaction takes place:

$$Cu(s) + Ag^+(aq) \rightarrow Cu^{2+}(aq) + Ag(s)$$

1. Represent the cell using standard notation.
2. Calculate the cell potential (emf) of the electrochemical cell.

Answer
Step 1 : Write equations for the two half reactions involved
$Cu^{2+} + 2e^- \Leftrightarrow Cu$ (E^{o}V = 0.16V)
$Ag^+ + e^- \Leftrightarrow Ag$ (E^{o}V = 0.80V)

Step 2 : Determine which reaction takes place at the cathode and which is the anode reaction
Both half-reactions have positive electrode potentials, but the silver half-reaction has a higher positive value. In other words, silver does not form ions easily, and this must be the reduction half-reaction. Copper is the oxidation half-reaction. Copper is oxidised, therefore this is the anode reaction. Silver is reduced and so this is the cathode reaction.
Step 3 : Represent the cell using standard notation

$$Cu|Cu^{2+}(1mol.dm^{-3})||Ag^+(1mol.dm^{-3})|Ag$$

Step 4 : Calculate the cell potential
E$^0_{(cell)}$ = E^0 (cathode) - E^0 (anode)
= +0.80 - (+0.34)
= +0.46 V

Worked Example 87: Calculating the emf of a cell

Question: Calculate the cell potential of the electrochemical cell in which the following reaction takes place, and represent the cell using standard notation.

$$Mg(s) + 2H^+(aq) \rightarrow Mg2+(aq) + H_2(g)$$

Answer
Step 1 : Write equations for the two half reactions involved
$Mg^{2+} + 2e^- \Leftrightarrow Mg$ (E^{o}V = -2.37)
$2H^+ + 2e^- \Leftrightarrow H_2$ (E^{o}V = 0.00)

Step 2 : Determine which reaction takes place at the cathode and which is the anode reaction
From the overall equation, it is clear that magnesium is oxidised and hydrogen ions are reduced in this reaction. Magnesium is therefore the anode reaction and hydrogen is the cathode reaction.
Step 3 : Represent the cell using standard notation

$$Mg|Mg^{2+}||H^+|H_2$$

Step 4 : Calculate the cell potential
E$^0_{(cell)}$ = E^0 (cathode) - E^0 (anode)
= 0.00 - (-2.37)
= +2.37 V

Predicting whether a reaction will take place spontaneously

Look at the following example to help you to understand how to predict whether a reaction will take place spontaneously or not.

In the reaction,

$$Pb^{2+}(aq) + 2Br^{-}(aq) \rightarrow Br_2(l) + Pb(s)$$

the two half reactions are as follows:

$$Pb^{2+} + 2e^{-} \Leftrightarrow Pb \text{ (-0.13 V)}$$
$$Br_2 + 2e^{-} \Leftrightarrow 2Br^{-} \text{ (+1.06 V)}$$

Important: Half cell reactions

You will see that the half reactions are written as they appear in the table of standard electrode potentials. It may be useful to highlight the reacting substance in each half reaction. In this case, the reactants are Pb^{2+} and Br^{-} ions.

Look at the electrode potential for the first half reaction. The negative value shows that lead loses electrons easily, in other words it is easily oxidised. The reaction would normally proceed from right to left (i.e. the equilibrium lies to the left), but in the original equation, the opposite is happening. It is the Pb^{2+} ions that are being reduced to lead. This part of the reaction is therefore not spontaneous. The positive electrode potential value for the bromine half-reaction shows that bromine is more easily reduced, in other words the equilibrium lies to the right. The spontaneous reaction proceeds from left to right. This is not what is happening in the original equation and therefore this is also not spontaneous. Overall it is clear then that the reaction will not proceed spontaneously.

Worked Example 88: Predicting whether a reaction is spontaneous

Question: Will copper react with dilute sulfuric acid (H_2SO_4)? You are given the following half reactions:

$$Cu^{2+}(aq) + 2e^{-} \Leftrightarrow Cu(s) \text{ (E}^0 = +0.34 \text{ V)}$$
$$2H^{+}(aq) + 2e^{-} \Leftrightarrow H_2(g) \text{ (E}^0 = 0 \text{ V)}$$

Answer

Step 5 : For each reaction, look at the electrode potentials and decide in which direction the equilibrium lies

In the first half reaction, the positive electrode potential means that copper does not lose electrons easily, in other words it is more easily reduced and the equilibrium position lies to the right. Another way of saying this is that the spontaneous reaction is the one that proceeds from left to right, when copper ions are reduced to copper metal.

In the second half reaction, the spontaneous reaction is from right to left.

Step 6 : Compare the equilibrium positions to the original reaction

What you should notice is that in the original reaction, the reactants are copper (Cu) and sulfuric acid ($2H^{+}$). During the reaction, the copper is oxidised and the hydrogen ions are reduced. But from an earlier step, we know that neither of these half reactions will proceed spontaneously in the direction indicated by the original reaction. The reaction is therefore not spontaneous.

Important:

A second method for predicting whether a reaction is spontaneous

Another way of predicting whether a reaction occurs spontaneously, is to look at the sign of the emf value for the cell. If the emf is **positive** then the reaction is **spontaneous**. If the emf is **negative**, then the reaction is **not spontaneous**.

Balancing redox reactions

We will look at this in more detail in the next section.

Exercise: Predicting whether a reaction will take place spontaneously

1. Predict whether the following reaction will take place spontaneously or not. Show all your working.

$$2Ag(s) + Cu^{2+}(aq) \rightarrow Cu(s) + 2Ag^{+}(aq)$$

2. Zinc metal reacts with an acid, H^+ (aq) to produce hydrogen gas.
   (a) Write an equation for the reaction, using the table of electrode potentials.
   (b) Predict whether the reaction will take place spontaneously. Show your working.
3. Four beakers are set up, each of which contains one of the following solutions:
   (a) $Mg(NO_3)_2$
   (b) $Ba(NO_3)_2$
   (c) $Cu(NO_3)_2$
   (d) $Al(NO_3)_2$

   Iron is added to each of the beakers. In which beaker will a spontaneous reaction take place?
4. Which one of the following solutions can be stored in an aluminium container?
   (a) $Cu(SO)_4$
   (b) $Zn(SO)_4$
   (c) NaCl
   (d) $Pb(NO_3)_2$

Exercise: Electrochemical cells and standard electrode potentials

1. An electrochemical cell is made up of a copper electrode in contact with a copper nitrate solution and an electrode made of an unknown metal M in contact with a solution of MNO_3. A salt bridge containing a KNO_3 solution joins the two half cells. A voltmeter is connected across the electrodes. Under standard conditions the reading on the voltmeter is 0.46V.

The reaction in the copper half cell is given by:
$\mathrm{Cu} \rightarrow \mathrm{Cu}^{2+} + 2\mathrm{e}^-$

(a) Write down the standard conditions which apply to this electrochemical cell.
(b) Identify the metal M. Show calculations.
(c) Use the standard electrode potentials to write down equations for the:
  i. cathode half-reaction
  ii. anode half-reaction
  iii. overall cell reaction
(d) What is the purpose of the salt bridge?
(e) Explain why a KCl solution would not be suitable for use in the salt bridge in this cell.

(IEB Paper 2, 2004)

2. Calculate the emf for each of the following standard electrochemical cells:

(a)
$$Mg|Mg^{2+}||H^+|H_2$$
(b)
$$Fe|Fe^{3+}||Fe^{2+}|Fe$$
(c)
$$Cr|Cr2+||Cu^{2+}|Cu$$
(d)
$$Pb|Pb^{2+}||Hg^{2+}|Hg$$

3. Given the following two half-reactions:
   - $Fe^{3+}(aq) + e^- \Leftrightarrow Fe^{2+}(aq)$
   - $MnO_4^-(aq) + 8H^+(aq) + 5e^- \Leftrightarrow Mn^{2+}(aq) + 4H_2O(l)$

(a) Give the standard electrode potential for each half-reaction.
(b) Which reaction takes place at the cathode and which reaction takes place at the anode?
(c) Represent the electrochemical cell using standard notation.
(d) Calculate the emf of the cell

17.5 Balancing redox reactions

Half reactions can be used to balance redox reactions. We are going to use some worked examples to help explain the method.

Worked Example 89: Balancing redox reactions

Question: Magnesium reduces copper (II) oxide to copper. In the process, magnesium is oxidised to magnesium ions. Write a balanced equation for this reaction.

Answer

Step 1 : Write down the unbalanced oxidation half reaction.

$$Mg \rightarrow Mg^{2+}$$

Step 2 : Balance the number of atoms on both sides of the equation.
You are allowed to add hydrogen ions (H^+) and water molecules if the reaction takes place in an acid medium. If the reaction takes place in a basic medium, you can add either hydroxide ions (OH^-) or water molecules. In this case, there is one magnesium atom on the left and one on the right, so no additional atoms need to be added.

Step 3 : Once the atoms are balanced, check that the charges balance.
Charges can be balanced by adding electrons to either side. The charge on the left of the equation is 0, but the charge on the right is +2. Therefore, two electrons must be added to the right hand side so that the charges balance. The half reaction is now:

$$Mg \rightarrow Mg^{2+} + 2e^-$$

Step 4 : Repeat the above steps, but this time using the reduction half reaction.
The reduction half reaction is:

$$Cu^{2+} \rightarrow Cu$$

The atoms balance but the charges don't. Two electrons must be added to the right hand side.

$$Cu^{2+} + 2e^- \rightarrow Cu$$

Step 5 : Multiply each half reaction by a suitable number so that the number of electrons released in the oxidation half reaction is made equal to the number of electrons that are accepted in the reduction half reaction.
No multiplication is needed because there are two electrons on either side.

Step 6 : Combine the two half reactions to get a final equation for the overall reaction.

$$Mg + Cu^{2+} + 2e^- \rightarrow Mg^{2+} + Cu + 2e^- \text{ (The electrons on either side cancel and you get...)}$$
$$Mg + Cu^{2+} \rightarrow Mg^{2+} + Cu$$

Step 7 : Do a final check to make sure that the equation is balanced
In this case, it is.

Worked Example 90: Balancing redox reactions

Question: Chlorine gas oxidises Fe(II) ions to Fe(III) ions. In the process, chlorine is reduced to chloride ions. Write a balanced equation for this reaction.

Answer
Step 1 : Write down the oxidation half reaction.

$$Fe^{2+} \rightarrow Fe^{3+}$$

Step 2 : Balance the number of atoms on both sides of the equation.
There is one iron atom on the left and one on the right, so no additional atoms need to be added.

Step 3 : Once the atoms are balanced, check that the charges balance.
The charge on the left of the equation is +2, but the charge on the right is +3. Therefore, one electron must be added to the right hand side so that the charges balance. The half reaction is now:

$$Fe^{2+} \rightarrow Fe^{3+} + e^-$$

Step 4 : Repeat the above steps, but this time using the reduction half reaction.
The reduction half reaction is:

$$Cl_2 \rightarrow Cl^-$$

The atoms don't balance, so we need to multiply the right hand side by two to fix this. Two electrons must be added to the left hand side to balance the charges.

$$Cl_2 + 2e^- \rightarrow 2Cl^-$$

Step 5 : Multiply each half reaction by a suitable number so that the number of electrons released in the oxidation half reaction is made equal to the number of electrons that are accepted in the reduction half reaction.
We need to multiply the oxidation half reaction by two so that the number of electrons on either side are balanced. This gives:

$$2Fe^{2+} \rightarrow 2Fe^{3+} + 2e^-$$

Step 6 : Combine the two half reactions to get a final equation for the overall reaction.

$$2Fe^{2+} + Cl_2 \rightarrow 2Fe^{3+} + 2Cl^-$$

Step 7 : Do a final check to make sure that the equation is balanced
The equation is balanced.

Worked Example 91: Balancing redox reactions in an acid medium

Question: The following reaction takes place in an acid medium:

$$Cr_2O_7^{2-} + H_2S \rightarrow Cr^{3+} + S$$

Write a balanced equation for this reaction.

Answer
Step 1 : Write down the oxidation half reaction.

$$Cr_2O_7^{2-} \rightarrow Cr^{3+}$$

Step 2 : Balance the number of atoms on both sides of the equation.
We need to multiply the right side by two so that the number of Cr atoms will balance. To balance the oxygen atoms, we will need to add water molecules to the right hand side.

$$Cr_2O_7^{2-} \rightarrow 2Cr^{3+} + 7H_2O$$

Now the oxygen atoms balance but the hydrogens don't. Because the reaction takes place in an acid medium, we can add hydrogen ions to the left side.

$$Cr_2O_7^{2-} + 14H^+ \rightarrow 2Cr^{3+} + 7H_2O$$

Step 3 : Once the atoms are balanced, check that the charges balance.
The charge on the left of the equation is (-2+14) = +12, but the charge on the right is +6. Therefore, six electrons must be added to the left hand side so that the charges balance. The half reaction is now:

$$Cr_2O_7^{2-} + 14H^+ + 6e^- \rightarrow 2Cr^{3+} + 7H_2O$$

Step 4 : Repeat the above steps, but this time using the reduction half reaction.
The reduction half reaction after the charges have been balanced is:

$$S^{2-} \rightarrow S + 2e^-$$

Step 5 : Multiply each half reaction by a suitable number so that the number of electrons released in the oxidation half reaction is made equal to the number of electrons that are accepted in the reduction half reaction.
We need to multiply the reduction half reaction by three so that the number of electrons on either side are balanced. This gives:

$$3S^{2-} \rightarrow 3S + 6e^-$$

Step 6 : Combine the two half reactions to get a final equation for the overall reaction.

$$Cr_2O_7^{2-} + 14H^+ + 3S^{2-} \rightarrow 3S + 2Cr^{3+} + 7H_2O$$

Step 7 : Do a final check to make sure that the equation is balanced

Worked Example 92: Balancing redox reactions in an alkaline medium

Question: If ammonia solution is added to a solution that contains cobalt(II) ions, a complex ion is formed, called the hexaaminecobalt(II) ion ($Co(NH_3)_6^{2+}$). In a chemical reaction with hydrogen peroxide solution, hexaaminecobalt ions are oxidised by hydrogen peroxide solution to the hexaaminecobalt(III) ion $Co(NH_3)_6^{3+}$. Write a balanced equation for this reaction.

Answer
Step 1 : Write down the oxidation half reaction

$$Co(NH_3)_6^{2+} \rightarrow Co(NH_3)_6^{3+}$$

Step 2 : Balance the number of atoms on both sides of the equation.
The number of atoms are the same on both sides.

Step 3 : Once the atoms are balanced, check that the charges balance.
The charge on the left of the equation is +2, but the charge on the right is +3. One elctron must be added to the right hand side to balance the charges in the equation.The half reaction is now:

$$Co(NH_3)_6^{2+} \rightarrow Co(NH_3)_6^{3+} + e^-$$

Step 4 : Repeat the above steps, but this time using the reduction half reaction.
Although you don't actually know what product is formed when hydrogen peroxide is reduced, the most logical product is OH^-. The reduction half reaction is:

$$H_2O_2 \rightarrow OH^-$$

After the atoms and charges have been balanced, the final equation for the reduction half reaction is:

$$H_2O_2 + 2e^- \rightarrow 2OH^-$$

Step 5 : Multiply each half reaction by a suitable number so that the number of electrons released in the oxidation half reaction is made equal to the number of electrons that are accepted in the reduction half reaction.
We need to multiply the oxidation half reaction by two so that the number of electrons on both sides are balanced. This gives:

$$2Co(NH_3)_6^{2+} \rightarrow 2Co(NH_3)_6^{3+} + 2e^-$$

Step 6 : Combine the two half reactions to get a final equation for the overall reaction.

$$2Co(NH_3)_6^{2+} + H_2O_2 \rightarrow 2Co(NH_3)_6^{3+} + 2OH^-$$

Step 7 : Do a final check to make sure that the equation is balanced

Exercise: Balancing redox reactions

1. Balance the following equations.
   (a) $HNO_3 + PbS \rightarrow PbSO_4 + NO + H_2O$
   (b) $NaI + Fe_2(SO_4)_3 \rightarrow I_2 + FeSO_4 + Na_2SO_4$
2. Manganate(VII) ions (MnO_4^-) oxidise hydrogen peroxide (H_2O_2) to oxygen gas. The reaction is done in an acid medium. During the reaction, the manganate(VII) ions are reduced to manganese(II) ions (Mn^{2+}). Write a balanced equation for the reaction.
3. Chlorine gas is prepared in the laboratory by adding concentrated hydrochloric acid to manganese dioxide powder. The mixture is carefully heated.
   (a) Write down a balanced equation for the reaction which takes place.
   (b) Using standard electrode potentials, show by calculations why this mixture needs to be heated.
   (c) Besides chlorine gas which is formed during the reaction, hydrogen chloride gas is given off when the conentrated hydrochloric acid is heated. Explain why the hydrogen chloride gas is removed from the gas mixture when the gas is bubbled through water.
   (IEB Paper 2, 2004)

4. The following equation can be deduced from the table of standard electrode potentials:

   $$2Cr_2O_7^{2-}(aq) + 16H^+(aq) \rightarrow 4Cr^{3+}(aq) + 3O_2(g) + 8H_2O(l) \text{ (E}^0 = +0.10\text{V)}$$

   This equation implies that an acidified solution of aqueous potassium dichromate (orange) should react to form Cr^{3+} (green). Yet aqueous laboratory solutions of potassium dichromate remain orange for years. Which ONE of the following best explains this?

   (a) Laboratory solutions of aqueous potassium dichromate are not acidified
   (b) The E^0 value for this reaction is only +0.10V
   (c) The activation energy is too low
   (d) The reaction is non-spontaneous

   (IEB Paper 2, 2002)

5. Sulfur dioxide gas can be prepared in the laboratory by heating a mixture of copper turnings and concentrated sulfuric acid in a suitable flask.

   (a) Derive a balanced ionic equation for this reaction using the half-reactions that take place.
   (b) Give the E^0 value for the overall reaction.
   (c) Explain why it is necessary to heat the reaction mixture.
   (d) The sulfur dioxide gas is now bubbled through an aqueous solution of potassium dichromate. Describe and explain what changes occur during this process.

   (IEB Paper 2, 2002)

17.6 Applications of electrochemistry

Electrochemistry has a number of different uses, particularly in industry. We are going to look at a few examples.

17.6.1 Electroplating

Electroplating is the process of using electrical current to coat an electrically conductive object with a thin layer of metal. Mostly, this application is used to deposit a layer of metal that has some desired property (e.g. abrasion and wear resistance, corrosion protection, improvement of aesthetic qualities etc.) onto a surface that doesn't have that property. Electro-refining (also sometimes called *electrowinning* is electroplating on a large scale. Electrochemical reactions are used to deposit pure metals from their ores. One example is the electrorefining of copper.

Copper plays a major role in the electrical reticulation industry as it is very conductive and is used in electric cables. One of the problems though is that copper must be pure if it is to be an effective current carrier. One of the methods used to purify copper, is electro-winning. The copper electro-winning process is as follows:

1. Bars of crude (impure) copper containing other metallic impurities is placed on the *anodes*.

2. The *cathodes* are made up of pure copper with few impurities.

3. The electrolyte is a solution of aqueous $CuSO_4$ and H_2SO_4.

4. When current passes through the cell, electrolysis takes place. The impure copper anode dissolves to form Cu^{2+} ions in solution. These positive ions are attracted to the negative cathode, where reduction takes place to produce pure copper metal. The reactions that take place are as follows:

   At the *anode*:
   $$Cu(s) \rightarrow Cu^{2+}(aq) + 2e^-$$

   At the *cathode*:
   $$Cu^{+2}(aq) + 2e^- \rightarrow Cu(s) \qquad (> 99\%\text{purity})$$

5. The other metal impurities (Zn, Au, Ag, Fe and Pb) do not dissolve and form a solid sludge at the bottom of the tank or remain in solution in the electrolyte.

Figure 17.6: A simplified diagram to illustrate what happens during the electrowinning of copper

17.6.2 The production of chlorine

Electrolysis can also be used to produce chlorine gas from brine/seawater (NaCl). This is sometimes referred to as the 'Chlor-alkali' process. The reactions that take place are as follows:

At the *anode* the reaction is:

$$2Cl^- \rightarrow Cl_2(g) + 2e^-$$

whereas at the *cathode*, the following happens:

$$2Na^+ + 2H_2O + 2e^- \rightarrow 2Na^+ + 2OH^- + H_2$$

The overall reaction is:

$$2Na^+ + 2H_2O + 2Cl^- \rightarrow 2Na^+ + 2OH^- + H_2 + Cl_2$$

Chlorine is a very important chemical. It is used as a bleaching agent, a disinfectant, in solvents, pharmaceuticals, dyes and even plastics such as polyvinlychloride (PVC).

Figure 17.7: The electrolysis of sodium chloride

17.6.3 Extraction of aluminium

Aluminum metal is a commonly used metal in industry where its properties of being both light and strong can be utilized. It is also used in the manufacture of products such as aeroplanes and motor cars. The metal is present in deposits of bauxite which is a mixture of silicas, iron oxides and hydrated alumina (Al_2O_3 x H_2O).

Electrolysis can be used to extract aluminum from bauxite. The process described below produces 99% pure aluminum:

1. Aluminum is melted along with cryolite (Na_3AlF_6) which acts as the electrolyte. Cryolite helps to lower the melting point and dissolve the ore.
2. The *anode* carbon rods provide sites for the oxidation of O^{2-} and F^- ions. Oxygen and flourine gas are given off at the anodes and also lead to anode consumption.
3. At the *cathode* cell lining, the Al^{3+} ions are reduced and metal aluminum deposits on the lining.
4. The AlF_6^{3-} electrolyte is stable and remains in its molten state.

The basic electrolytic reactions involved are as follows: At the *cathode*:

$$Al^{+3} + 3e^- \quad \rightarrow \quad Al(s) \qquad (99\%\text{purity})$$

At the *anode*:

$$2O^{2-} \quad \rightarrow \quad O_2(g) + 4e-$$

The overall reaction is as follows:

$$2Al_2O_3 \quad \rightarrow \quad 4Al + 3O_2$$

The only problem with this process is that the reaction is *endothermic* and large amounts of electricity are needed to drive the reaction. The process is therefore very expensive.

17.7 Summary

- An **electrochemical reaction** is one where either a chemical reaction produces an external voltage, or where an external voltage causes a chemical reaction to take place.

- In a **galvanic cell** a chemical reaction produces a current in the external circuit. An example is the zinc-copper cell.
- A galvanic cell has a number of **components**. It consists of two **electrodes**, each of which is placed in a separate beaker in an **electrolyte** solution. The two electrolytes are connected by a **salt bridge**. The electrodes are connected two each other by an external circuit wire.
- One of the electrodes is the **anode**, where **oxidation** takes place. The **cathode** is the electrode where **reduction** takes place.
- In a galvanic cell, the build up of electrons at the anode sets up a potential difference between the two electrodes, and this causes a **current** to flow in the external circuit.
- A **galvanic cell** is therefore an electrochemical cell that uses a chemical reaction between two dissimilar electrodes dipped in an electrolyte to generate an electric current.
- The standard notation for a galvanic cell such as the zinc-copper cell is as follows:

$$Zn|Zn^{2+}||Cu^{2+}|Cu$$

  where

$$\begin{aligned} | &= \text{a phase boundary (solid/aqueous)} \\ || &= \text{the salt bridge} \end{aligned}$$

- The galvanic cell is used in **batteries** and in **electroplating**.
- An **electrolytic cell** is an electrochemical cell that uses electricity to drive a non-spontaneous reaction. In an electrolytic cell, **electrolysis** occurs, which is a process of separating elements and compounds using an electric current.
- One example of an electrolytic cell is the electrolysis of copper sulphate to produce copper and sulphate ions.
- Different metals have different **reaction potentials**. The reaction potential of metals (in other words, their ability to ionise), is recorded in a **standard table of electrode potential**. The more negative the value, the greater the tendency of the metal to be oxidised. The more positive the value, the greater the tendency of the metal to be reduced.
- The values on the standard table of electrode potentials are measured relative to the **standard hydrogen electrode**.
- The emf of a cell can be calculated using one of the following equations:

  $E^0_{(cell)}$ = E^0 (right) - E^0 (left)

  $E^0_{(cell)}$ = E^0 (reduction half reaction) - E^0 (oxidation half reaction)

  $E^0_{(cell)}$ = E^0 (oxidising agent) - E^0 (reducing agent)

  $E^0_{(cell)}$ = E^0 (cathode) - E^0 (anode)
- It is possible to predict whether a reaction is **spontaneous** or not, either by looking at the sign of the cell's emf or by comparing the electrode potentials of the two half cells.
- It is possible to **balance** redox equations using the half-reactions that take place.
- There are a number of important **applications** of electrochemistry. These include **electroplating**, the **production of chlorine** and the **extraction of aluminium**.

Exercise: Summary exercise

1. For each of the following, say whether the statement is **true** or **false**. If it is false, re-write the statement correctly.
   (a) The anode in an electrolytic cell has a negative charge.
   (b) The reaction $2\mathrm{KClO_3} \rightarrow 2\mathrm{KCl} + 3\mathrm{O_2}$ is an example of a redox reaction.
   (c) Lead is a stronger oxidising agent than nickel.
2. For each of the following questions, choose the **one correct** answer.
   (a) Which one of the following reactions is a redox reaction?
      i. $HCl + NaOH \rightarrow NaCl + H_2O$
      ii. $AgNO_3 + NaI \rightarrow AgI + NaNO_3$
      iii. $2FeCl_3 + 2H_2O + SO_2 \rightarrow H_2SO_4 + 2HCl + 2FeCl_2$
      iv. $BaCl_2 + MgSO_4 \rightarrow MgCl_2 + BaSO_4$
      (IEB Paper 2, 2003)
   (b) Consider the reaction represented by the following equation:
      $Br_{2(l)} + 2I^-_{aq} \rightarrow 2Br^-_{aq} + I_{2(s)}$
      Which one of the following statements about this reaction is correct?
      i. bromine is oxidised
      ii. bromine acts as a reducing agent
      iii. the iodide ions are oxidised
      iv. iodine acts as a reducing agent
      (IEB Paper 2, 2002)
   (c) The following equations represent two hypothetical half-reactions:
      $X_2 + 2e^- \Leftrightarrow 2X^-$ (+1.09 V) and
      $Y^+ + e^- \Leftrightarrow Y$ (-2.80 V)
      Which one of the following substances from these half-reactions has the greatest tendency to donate electrons?
      i. X^-
      ii. X_2
      iii. Y
      iv. Y^+
   (d) Which one of the following redox reactions will **not** occur spontaneously at room temperature?
      i. $Mn + Cu^{2+} \rightarrow Mn^{2+} + Cu$
      ii. $Zn + SO_4^{2-} + 4H^+ \rightarrow Zn^{2+} + SO_2 + 2H_2O$
      iii. $Fe^{3+} + 3NO_2 + 3H_2O \rightarrow Fe + 3NO_3^- + 6H^+$
      iv. $5H_2S + 2MnO_4^- + 6H^+ \rightarrow 5S + 2Mn^{2+} + 8H_2O$
   (e) Which statement is CORRECT for a Zn-Cu galvanic cell that operates under standard conditions?
      i. The concentration of the Zn^{2+} ions in the zinc half-cell gradually decreases.
      ii. The concentration of the Cu^{2+} ions in the copper half-cell gradually increases.
      iii. Negative ions migrate from the zinc half-cell to the copper half-cell.
      iv. The intensity of the colour of the electrolyte in the copper half-cell gradually decreases.
      (DoE Exemplar Paper 2, 2008)
3. In order to investigate the rate at which a reaction proceeds, a learner places a beaker containing concentrated nitric acid on a sensitive balance. A few pieces of copper metal are dropped into the nitric acid.
   (a) Use the relevant half-reactions from the table of Standard Reduction Potentials to derive the balanced nett ionic equation for the reaction that takes place in the beaker.
   (b) What chemical property of nitric acid is illustrated by this reaction?
   (c) List three observations that this learner would make during the investigation.

(IEB Paper 2, 2005)

4. The following reaction takes place in an electrochemical cell:

$$Cu(s) + 2AgNO_3(aq) \rightarrow Cu(NO_3)_2(aq) + 2Ag(s)$$

   (a) Give an equation for the oxidation half reaction.
   (b) Which metal is used as the anode?
   (c) Determine the emf of the cell under standard conditions.

   (IEB Paper 2, 2003)

5. The nickel-cadmium (NiCad) battery is small and light and is made in a sealed unit. It is used in portable appliances such as calculators and electric razors. The following two half reactions occur when electrical energy is produced by the cell.
   Half reaction 1: $\mathrm{Cd(s) + 2OH^-(aq) \rightarrow Cd(OH)_2(s) + 2e^-}$
   Half reaction 2: $\mathrm{NiO(OH)(s) + H_2O(l) + e^- \rightarrow Ni(OH)_2(s) + OH^-(aq)}$
   (a) Which half reaction (1 or 2) occurs at the anode? Give a reason for your answer.
   (b) Which substance is oxidised?
   (c) Derive a balanced ionic equation for the overall cell reaction for the discharging process.
   (d) Use your result above to state in which direction the cell reaction will proceed (forward or reverse) when the cell is being charged.

   (IEB Paper 2, 2001)

6. An electrochemical cell is constructed by placing a lead rod in a porous pot containing a solution of lead nitrate (see sketch). The porous pot is then placed in a large aluminium container filled with a solution of aluminium sulphate. The lead rod is then connected to the aluminium container by a copper wire and voltmeter as shown.

   (a) Define the term *reduction*.
   (b) In which direction do electrons flow in the copper wire? (Al to Pb or Pb to Al)
   (c) Write balanced equations for the reactions that take place at...
      i. the anode
      ii. the cathode
   (d) Write a balanced nett ionic equation for the reaction which takes place in this cell.
   (e) What are the two functions of the porous pot?
   (f) Calculate the emf of this cell under standard conditions.

   (IEB Paper 2, 2005)

Part IV

Chemical Systems

Chapter 18

The Water Cycle - Grade 10

18.1 Introduction

You may have heard the word 'cycle' many times before. Think for example of the word 'bicycle' or the regular 'cycle tests' that you may have at school. A cycle is a series of events that repeats itself. In the case of a bicycle, the wheel turns through a full circle before beginning the motion again, while cycle tests happen regularly, normally every week or every two weeks. Because a cycle repeats itself, it doesn't have a beginning or an end.

Our Earth is a **closed system**. This means that it can exchange *energy* with its surroundings (i.e. the rest of the solar system), but no new *matter* is brought into the system. For this reason, it is important that all the elements and molecules on Earth are recycled so that they are never completely used up. In the next two sections, we are going to take a closer look at two cycles that are very important for life on Earth. They are the **water cycle** and the **nitrogen cycle**.

18.2 The importance of water

For many people, it is so easy to take water for granted, and yet life on Earth would not exist were it not for this extraordinary compound. Not only is it believed that the first forms of life actually *started* in water, but most of the cells in living organisms contain between 70% and 95% water. Here in the cells, water acts as a solvent and helps to transport vital materials such as food and oxygen to where they are needed, and also removes waste products such as carbon dioxide and ammonia from the body. For many animals and plants, water is their home. Think for example of fish and amphibians that live either all or part of the time in rivers, dams and the oceans. In other words, if water did not exist, no life would be possible.

Apart from allowing life to exist, water also has a number of other functions. Water shapes the landscape around us by wearing away at rocks and also transports and deposits sediments on floodplains and along coastal regions. Water also plays a very important role in helping to regulate Earth's climate. We will discuss this again later in the chapter. As humans we use water in our homes, in industry, in mining, irrigation and even as a source of electricitiy in hydro-electric schemes. In fact, if we were able to view Earth from space, we would see that almost three quarters of our planet's surface is covered in water. It is because of this that Earth is sometimes called the 'Blue Planet'. Most of this water is stored in the oceans, with the rest found in ice (e.g. glaciers), groundwater (e.g. boreholes), surface water (e.g. rivers, lakes, estuaries, dams) and in the atmosphere as clouds and water vapour.

Interesting Fact

In the search for life on other planets, one of the first things that scientists look for is water. However, most planets are either too close to the sun (and therefore

too hot) for water to exist in liquid form, or they are too far away and therefore too cold. So, even if water were to be found, the conditions are unlikely to allow it to exist in a form that can support the diversity of life that we see on Earth.

18.3 The movement of water through the water cycle

The water cycle is the continuous movement of water over, above, and beneath the Earth's surface. As water moves, it changes phase between liquid (water), solid (ice) and gas (water vapour). It is powered by solar energy and, because it is a cycle, it has no beginning or end.

Definition: The Water Cycle
The water cycle is the continuous circulation of water across the Earth. The water cycle is driven by solar radiation and it includes the atmosphere, land, surface water and groundwater. As water moves through the cycle, it changes state between liquid, solid, and gas phases. The actual movement of water from one part of the cycle to another (e.g. from river to ocean) is the result of processes such as evaporation, precipitation, infiltration and runoff.

The movement of water through the water cycle is shown in figure 18.1. In the figure, each process within this cycle is numbered. Each process will be described below.

1. **The source of energy**

   The water cycle is driven by the sun, which provides the heat energy that is needed for many of the other processes to take place.

2. **Evaporation**

   When water on the earth's surface is heated by the sun, the average energy of the water molecules increases and some of the molecules are able to leave the liquid phase and become water vapour. This is called evaporation. Evaporation is the change of water from a liquid to a gas as it moves from the ground, or from bodies of water like the ocean, rivers and dams, into the atmosphere.

3. **Transpiration**

   Transpiration is the evaporation of water from the aerial parts of plants, especially the leaves but also from the stems, flowers and fruits. This is another way that liquid water can enter the atmosphere as a gas.

4. **Condensation**

   When evaporation takes place, water vapour rises in the atmosphere and cools as the altitude (height above the ground) increases. As the temperature drops, the energy of the water vapour molecules also decreases, until the molecules don't have enough energy to stay in the gas phase. At this point, condensation occurs. Condensation is the change of water from water vapour (gas) into liquid water droplets in the air. *Clouds*, *fog* and *mist* are all examples of condensation. A cloud is actually a collection of lots and lots of tiny water droplets. This mostly takes place in the upper atmosphere but can also take place close to the ground if there is a significant temperature change.

Interesting Fact

Have you ever tried breathing out on a very cold day? It looks as though you are breathing out smoke! The moist air that you breathe out is much warmer than the air outside your body. As this warm, moist air comes into

Figure 18.1: The water cycle

contact with the colder air outside, its temperature drops very quickly and the water vapour in the air you breathe out condenses. The 'smoke' that you see is actually formed in much the same way as clouds form in the upper atmosphere.

5. **Precipitation**

   Precipitation occurs when water falls back to the earth's surface in the form of rain or snow. Rain will fall as soon as a cloud becomes too saturated with water droplets. Snow is similar to rain, except that it is frozen. Snow only falls if temperatures in the atmosphere are around freezing. The freeing point of water is 0^0C).

6. **Infiltration**

   If precipitation occurs, some of this water will filter into the soil and collect underground. This is called infiltration. This water may evaporate again from the soil at a later stage, or the underground water may seep into another water body.

7. **Surface runoff**

   This refers to the many ways that water moves across the land. This includes *surface runoff* such as when water flows along a road and into a drain, or when water flows straight across the sand. It also includes *channel runoff* when water flows in rivers and streams. As it flows, the water may infiltrate into the ground, evaporate into the air, become stored in lakes or reservoirs, or be extracted for agricultural or other human uses.

Important: It is important to realise that the water cycle is all about **energy exchanges**. The sun is the original energy *source*. Energy from the sun heats the water and causes evaporation. This energy is stored in water vapour as *latent heat*. When the water vapour condenses again, the latent heat is released, and helps to drive circulation in the atmosphere. The liquid water falls to earth, and will evaporate again at a later stage. The atmospheric circulation patterns that occur because of these exchanges of heat are very important in influencing climate patterns.

Activity :: Experiment : The Water Cycle

Materials:

Tile or piece of plastic (e.g. lid of ice-cream container) to make a hill slope; glass fish tank with a lid; beaker with ice cubes; lamp; water

Set up a model of the water cycle as follows:

1. Lean the plastic against one side so that it creates a 'hill slope' as shown in the diagram.
2. Pour water into the bottom of the tank until about a quarter of the hill slope is covered.
3. Close the fish tank lid.
4. Place the beaker with ice on the lid directly above the hill slope.
5. Turn the lamp on and position it so that it shines over the water.
6. Leave the model like this for 20-30 minutes and then observe what happens. Make sure that you don't touch the lamp as it will be very hot!

Observation questions:

1. Which parts of the water cycle can you see taking place in the model?
2. Which parts of the water cycle are *not* represented in the model?

3. Can you think of how those parts that are not shown *could* be represented?
4. What is the energy source in the model? What would the energy source be in reality?
5. What do you think the function of the ice is in the beaker?

18.4 The microscopic structure of water

In many ways, water behaves very differently from other liquids. These properties are directly related to the microscopic structure of water, and more specifically to the *shape* of the molecule and its *polar nature*, and to the *bonds* that hold water molecules together.

18.4.1 The polar nature of water

Every water molecule is made up of one oxygen atom that is bonded to two hydrogen atoms. When atoms bond, the nucleus of each atom has an attractive force on the electrons of the other atoms. This 'pull' is stronger in some atoms than in others and is called the **electronegativity** of the atom. In a water molecule, the oxygen atom has a higher electronegativty than the hydrogen atoms and therefore attracts the electrons more strongly. The result is that the oxygen atom has a slightly negative charge and the two hydrogen atoms each have a slightly positive charge. The water molecule is said to be **polar** because the electrical charge is not evenly distributed in the molecule. One part of the molecule has a different charge to other parts. You will learn more about this in chapter 4.

Figure 18.2: Diagrams showing the structure of a water molecule. Each molecule is made up of two hydrogen atoms that are attached to one oxygen atom.

18.4.2 Hydrogen bonding in water molecules

In every water molecule, the forces that hold the individual atoms together are called **intramolecular forces**. But there are also forces *between* different water molecules. These are called **intermolecular forces** (figure 18.3). You will learn more about these at a later stage, but for now it is enough to know that in water, molecules are held together by **hydrogen bonds**. Hydrogen bonds are a much stronger type of intermolecular force than those found in many other substances, and this affects the properties of water.

Important: Intramolecular and intermolecular forces

If you find these terms confusing, remember that 'intra' means *within* (i.e. the forces within a molecule). An **intro***vert* is someone who doesn't express emotions and feelings outwardly. They tend to be quieter and keep to themselves. 'Inter' means *between* (i.e. the forces between molecules). An **inter***national* cricket match is a match between two different countries.

Figure 18.3: Intermolecular and intramolecular forces in water. Note that the diagram on the left only shows *intermolecular* forces. The intramolecular forces are between the atoms of each water molecule.

18.5 The unique properties of water

Because of its polar nature and the strong hydrogen bonds between its molecules, water has some special properties that are quite different to those of other substances.

1. **Absorption of infra-red radiation**

   The polar nature of the water molecule means that it is able to absorb infra-red radiation (heat) from the sun. As a result of this, the oceans and other water bodies act as heat reservoirs, and are able to help moderate the Earth's climate.

2. **Specific heat**

Definition: Specific heat
Specific heat is the amount of heat energy that is needed to increase the temperature of a substance by one degree.

Water has a high specific heat, meaning that a lot of energy must be absorbed by water before its temperature changes.

Activity :: Demonstration : The high specific heat of water

(a) Pour about 100 ml of water into a glass beaker.

(b) Place the beaker on a stand and heat it over a bunsen burner for about 2 minutes.

(c) After this time, carefully touch the side of the beaker (Make sure you touch the glass very lightly because it will be very hot and may burn you!). Then use the end of a finger to test the temperature of the water.

What do you notice? Which of the two (glass or water) is the hottest?

You have probably observed this phenomenon if you have boiled water in a pot on the stove. The metal of the pot heats up very quickly, and can burn your fingers if you touch it, while the water may take several minutes before its temperature increases even slightly. How can we explain this in terms of hydrogen bonding? Remember that increasing the temperature of a substance means that its particles will move more quickly. However, before they can move faster, the bonds between them must be broken. In the case of water, these bonds are strong hydrogen bonds, and so a lot of energy is needed just to break these, before the particles can start moving faster.

It is the high specific heat of water and its ability to absorb infra-red radiation that allows it to regulate climate. Have you noticed how places that are closer to the sea have less extreme daily temperatures than those that are inland? During the day, the oceans heat up slowly, and so the air moving from the oceans across land is cool. Land temperatures are cooler than they would be if they were further from the sea. At night, the oceans lose the heat that they have absorbed very slowly, and so sea breezes blowing across the land are relatively warm. This means that at night, coastal regions are generally slightly warmer than areas that are further from the sea.

By contrast, places further from the sea experience higher maximum temperatures, and lower minimum temperatures. In other words, their *temperature range* is higher than that for coastal regions. The same principle also applies on a *global* scale. The large amount of water across Earth's surface helps to regulate temperatures by storing infra-red radiation (heat) from the sun, and then releasing it very slowly so that it never becomes too hot or too cold, and life is able to exist comfortably. In a similar way, water also helps to keep the temperature of the *internal* environment of living organisms relatively constant. This is very important. In humans, for example, a change in body temperature of only a few degrees can be deadly.

3. **Melting point and boiling point**

   The melting point of water is 0^0C and its boiling point is 100^0C. This large difference between the melting and boiling point is very important because it means that water can exist as a liquid over a large range of temperatures. The three phases of water are shown in figure 18.4.

4. **High heat of vaporisation**

Definition: Heat of vaporisation
Heat of vaporisation is the energy that is needed to change a given quantity of a substance into a gas.

The strength of the hydrogen bonds between water molecules also means that it has a high heat of vaporisation. 'Heat of vaporisation' is the heat energy that is needed to change water from the liquid to the gas phase. Because the bonds between molecules are strong, water has to be heated to 100^0C before it changes phase. At this temperature, the molecules have enough energy to break the bonds that hold the molecules together. The heat of vaporisation for water is 40.65 kJ/mol. It is very lucky for life on earth that water does have a high heat of vaporisation. Can you imagine what a problem it would be if water's heat of vaporisation was much lower? All the water that makes up the cells in our bodies would evaporate and most of the water on earth would no longer be able to exist as a liquid!

5. **Less dense solid phase**

   Another unusual property of water is that its solid phase (ice) is *less dense* than its liquid phase. You can observe this if you put ice into a glass of water. The ice doesn't sink to

Figure 18.4: Changes in phase of water

the bottom of the glass, but floats on top of the liquid. This phenomenon is also related to the hydrogen bonds between water molecules. While other materials contract when they solidify, water expands. The ability of ice to float as it solidifies is a very important factor in the environment. If ice sank, then eventually all ponds, lakes, and even the oceans would freeze solid as soon as temperatures dropped below freezing, making life as we know it impossible on Earth. During summer, only the upper few inches of the ocean would thaw. Instead, when a deep body of water cools, the floating ice insulates the liquid water below, preventing it from freezing and allowing life to exist under the frozen surface.

Figure 18.5: Ice cubes floating in water

Antarctica, the 'frozen continent', has one of the world's largest and deepest freshwater lakes. And this lake is hidden beneath 4 kilometres of ice! Lake Vostok is 200 km long and 50 km wide. The thick, glacial blanket of ice acts as an insulator, preventing the water from freezing.

6. **Water as a solvent**

Water is also a very good solvent, meaning that it is easy for other substances to dissolve in it. It is very seldom, in fact, that we find pure water. Most of the time, the water that we drink and use has all kinds of substances dissolved in it. It is these that make water taste different in different areas. So why, then, is it important that water is such a good solvent? We will look at just a few examples.

- Firstly, think about the animals and plants that live in aquatic environments such as rivers, dams or in the sea. All of these living organisms either need oxygen for respiration or carbon dioxide for photosynthesis, or both. How do they get these gases from the water in which they live? Oxygen and carbon dioxide are just two of the substances that dissolve easily in water, and this is how plants and animals obtain the gases that they need to survive. Instead of being available as gases in the atmosphere, they are present in solution in the surrounding water.
- Secondly, consider the fact that all plants need nitrogen to grow, and that they absorb this nitrogen from compounds such as nitrates and nitrates that are present in the soil. The question remains, however, as to how these nitrates and nitrites are able to be present in the soil at all, when most of the Earth's nitrogen is in a gaseous form in the atmosphere. Part of the answer lies in the fact that nitrogen oxides, which are formed during flashes of lightning, can be dissolved in rainwater and transported into the soil in this way, to be absorbed by plants. The other part of the answer lies in the activities of nitrogen-fixing bacteria in the soil, but this is a topic that we will return to in a later section.

It should be clear now, that water is an amazing compound, and that without its unique properties, life on Earth would definitely not be possible.

?

Exercise: The properties of water

1. A learner returns home from school on a hot afternoon. In order to get cold water to drink, she adds ice cubes to a glass of water. She makes the following observations:
   - The ice cubes float in the water.
   - After a while the water becomes cold and the ice cubes melt.

   (a) What property of ice cubes allows them to float in the water?
   (b) Briefly explain why the water gets cold when the ice cubes melt.
   (c) Briefly describe how the property you mentioned earlier affects the survival of aquatic life during winter.
2. Which properties of water allow it to remain in its liquid phase over a large temperature range? Explain why this is important for life on earth.

18.6 Water conservation

Water is a very precious substance and yet far too often, earth's water resources are abused and taken for granted. How many times have you walked past polluted rivers and streams, or seen the flow of water in a river reduced to almost nothing because of its extraction for industrial and other uses? And if you were able to test the *quality* of the water you see, you would probably be shocked. Often our water resources are contaminated with chemicals such as pesticides and fertilisers. If water is to continue playing all the important functions that were discussed earlier, it is vital that we reduce the impact of humans on these resources.

Activity :: Group work : Human impacts on the water cycle

Read the following extract from an article, entitled 'The Effects of Urbanisation on the Water Cycle' *by Susan Donaldson, and then answer the questions that follow.*

> As our communities grow, we notice many visible changes including housing developments, road networks, expansion of services and more. These changes have an impact on our precious water resources, with *pollution* of water being one of many such impacts. To understand these impacts you will need to have a good knowledge of the water cycle!
> It is interesting to note that the oceans contain most of earth's water (about 97%). Of the freshwater supplies on earth, 78% is tied up in polar ice caps and snow, leaving only a very small fraction available for use by humans. Of the available fresh water, 98% is present as groundwater, while the remaining 2% is in the form of surface water. Because our usable water supply is so limited, it is vitally important to protect water *quality*. Within the water cycle, there is no 'new' water ever produced on the earth. The water we use today has been in existence for billions of years. The water cycle continually renews and refreshes this finite water supply.
> So how exactly does urbanisation affect the water cycle? The increase in hard surfaces (e.g. roads, roofs, parking lots) decreases the amount of water that can soak into the ground. This increases the amount of surface runoff. The runoff water will collect many of the pollutants that have accumulated on these surfaces (e.g. oil from cars) and carry them into other water bodies such as rivers or the ocean. Because there is less infiltration, peak flows of stormwater runoff are larger and arrive earlier, increasing the size of urban floods. If groundwater supplies are reduced enough, this may affect stream flows during dry weather periods because it is the groundwater that seeps to the surface at these times.
> Atmospheric pollution can also have an impact because condensing water vapour will pick up these pollutants (e.g. SO_2, CO_2 and NO_2) and return them to earth into other water bodies. However, while the effects of urbanisation on water quality can be major, these impacts can be reduced if wise decisions are made during the process of development.

Questions

1. In groups, try to explain...
   (a) what is meant by 'urbanisation'
   (b) how urbanisation can affect water quality
2. Explain why it is so important to preserve the quality of our water supplies.
3. The article gives some examples of human impacts on water quality. In what *other* ways do human activities affect water quality?
4. What do you think some of the *consequences* of these impacts might be for humans and other forms of life?
5. Imagine that you are the city manager in your own city or the city closest to you. What changes would you introduce to try to protect the quality of water resources in your urban area?
6. What measures could be introduced in rural areas to protect water quality?

Apart from the *pollution* of water resources, the *overuse* of water is also a problem. In looking at the water cycle, it is easy sometimes to think that water is a never-ending resource. In a sense this is true because water cannot be destroyed. However, the *availability* of water may vary from place to place. In South Africa for example, many regions are extremely dry and receive very little rainfall. The same is true for many other parts of the world, where the scarcity of water

is a life and death issue. The present threat of **global warming** is also likely to affect water resources. Some climate models suggest that rising temperatures could increase the variability of climate and decrease rainfall in South Africa. With this in mind, and remembering that South Africa is already a dry country, it is vitally important that we manage our water use carefully. In addition to this, the less water there is available, the more likely it is that water *quality* will also decrease. A decrease in water quality limits how water can be used and developed.

At present, the demands being placed on South Africa's water resources are large. Table 18.1 shows the water requirements that were predicted for the year 2000. The figures in the table were taken from South Africa's *National Water Resource Strategy*, produced by the Department of Water Affairs and Forestry in 2004. In the table, 'rural' means water for domestic use and stock watering in rural areas, while 'urban' means water for domestic, industrial and commercial use in the urban area. 'Afforestation' is included because many plantations reduce stream flow because of the large amounts of water they need to survive.

Table 18.1: The predicted water requirements for various water management areas in South Africa for 2000 (million m^3/annum)

Water management area	Irrigation	Urban	Rural	Mining and bulk industrial	Power generation	Afforestation	Total
Limpopo	238	34	28	14	7	1	322
Thukela	204	52	31	46	1	0	334
Upper Vaal	114	635	43	173	80	0	1045
Upper Orange	780	126	60	2	0	0	968
Breede	577	39	11	0	0	6	633
Country total	**7920**	**2897**	**574**	**755**	**297**	**428**	**12871**

Activity :: Case Study : South Africa's water requirements
Refer to table 18.1 and then answer the following questions:

1. Which water management area in South Africa has the highest need for water...
   (a) in the *mining and industry* sector?
   (b) for *power generation*?
   (c) in the *irrigation* sector?
   (d) Suggest reasons for each of your answers above.
2. For South Africa as a whole...
   (a) Which activity uses the *most* water?
   (b) Which activity uses the *least* water?
3. Complete the following table, by calculating the percentage (%) that each activity contributes to the total water requirements in South Africa for the year 2000.

Water use activity	% of SA's total water requirements
Irrigation	
Urban	
Rural	
Mining and bulk industry	
Power generation	
Afforestation	

Table 18.2: The available water yield in South Africa in 2000 for various water management areas (million m^3/annum)

Water management area	Surface water	Ground	Irrigation	Urban	Mining and bulk industrial	Total local yield
Limpopo	160	98	8	15	0	281
Thukela	666	15	23	24	9	737
Upper Vaal	598	32	11	343	146	1130
Upper Orange	4311	65	34	37	0	4447
Breede	687	109	54	16	0	866
Country total	**10240**	**1088**	**675**	**970**	**254**	**13227**

Now look at table 18.2, which shows the amount of water *available* in South Africa during 2000. In the table, 'usable return flow' means the amount of water that can be reused after it has been used for irrigation, urban or mining.

Activity :: Case Study : Water conservation
Refer to table 18.2 and then answer the following questions:

1. Explain what is meant by...
   (a) surface water
   (b) ground water
2. Which water management area has the...
   (a) lowest surface water yield?
   (b) highest surface water yield?
   (c) lowest *total* yield?
   (d) highest *total* yield?
3. Look at the country's *total water requirements* for 2000 and the *total available yield*.
   (a) Calculate what percentage of the country's water yield is already being used up.
   (b) Do you think that the country's total water requirements will increase or decrease in the coming years? Give a reason for your answer.
4. South Africa is already placing a huge strain on existing water resources. In groups of 3-4, discuss ways that the country's demand for water could be reduced. Present your ideas to the rest of the class for discussion.

18.7 Summary

- Water is critical for the survival of life on Earth. It is an important part of the cells of living organisms and is used by humans in homes, industry, mining and agriculture.
- Water moves between the land and sky in the **water cycle**. The water cycle describes the changes in phase that take place in water as it circulates across the Earth. The water cycle is driven by solar radiation.

- Some of the important processes that form part of the water cycle are evaporation, transpiration, condensation, precipitation, infiltration and surface runoff. Together these processes ensure that water is cycled between the land and sky.
- It is the microscopic structure of water that determines its unique properties.
- Water molecules are **polar** and are held together by **hydrogen bonds**. These characteristics affect the properties of water.
- Some of the unique properties of water include its ability to absorb infra-red radiation, its high specific heat, high heat of vaporisation and the fact that the solid phase of water is less dense that its liquid phase.
- These properties of water help it to sustain life on Earth by moderating climate, regulating the internal environment of living organisms and allowing liquid water to exist below ice, even if temperatures are below zero.
- Water is also a good **solvent**. This property means that it is a good transport medium in the cells of living organisms, and that it can dissolve gases and other compounds that may be needed by aquatic plants and animals.
- Human activities threaten the quality of water resources through pollution and altered runoff patterns.
- As human populations grow, there is a greater demand for water. In many areas, this demand exceeds the amount of water available for use. Managing water wisely is important in ensuring that there will always be water available both for human use, and to maintain natural ecosystems.

Exercise: Summary Exercise

1. Give a word or term for each of the following phrases:
   (a) The continuous circulation of water across the earth.
   (b) The change in phase of water from gas to liquid.
   (c) The movement of water across a land surface.
   (d) The temperature at which water changes from liquid to gas.
2. In each of the following multiple choice questions, choose the one correct answer from the list provided.
   (a) Many of the unique properties of water (e.g. its high specific heat and high boiling point) are due to:
      i. strong covalent bonds between the hydrogen and oxygen atoms in each water molecule
      ii. the equal distribution of charge in a water molecule
      iii. strong hydrogen bonds between water molecules
      iv. the linear arrangement of atoms in a water molecule
   (b) Which of the following statements is **false**?
      i. Most of the water on earth is in the oceans.
      ii. The hardening of surfaces in urban areas results in increased surface runoff.
      iii. Water conservation is important because water cannot be recycled.
      iv. Irrigation is one of the largest water users in South Africa.
3. The sketch below shows a process that leads to rainfall in town X. The town has been relying only on rainfall for its water supply because it has no access to rivers or tap water. A group of people told the community that they will never run out of rainwater *because it will never stop raining*.

(a) List the processes labelled P_1 and P_2 that lead to rainfall in town X.

(b) Is this group of people correct in saying that town X will never run out of rainwater? Justify your answer using the sketch.
Recently, the amount of rainwater has decreased significantly. Various reasons have been given to explain the drought. Some of the community members are blaming this group who told them that it will never stop raining.

(c) What scientific arguments can you use to convince the community members that this group of people should not be blamed for the drought?

(d) What possible strategies can the community leaders adopt to ensure that they have a regular supply of water.

Chapter 19

Global Cycles: The Nitrogen Cycle - Grade 10

19.1 Introduction

The earth's atmosphere is made up of about 78% nitrogen, making it the largest pool of this gas. Nitrogen is essential for many biological processes. It is in all amino acids, proteins and nucleic acids. As you will see in a later chapter, these compounds are needed to build tissues, transport substances around the body, and control what happens in living organisms. In plants, much of the nitrogen is used in chlorophyll molecules which are needed for photosynthesis and growth.

So, if nitrogen is so essential for life, how does it go from being a gas in the atmosphere to being part of living organisms such as plants and animals? The problem with nitrogen is that it is an 'inert' gas, which means that it is unavailable to living organisms in its gaseous form. This is because of the strong triple bond between its atoms that makes it difficult to break. Something needs to happen to the nitrogen gas to change it into a form that it can be used. And at some later stage, these new compounds must be converted back into nitrogen gas so that the amount of nitrogen in the atmosphere stays the same. This process of changing nitrogen into different forms is called the **nitrogen cycle** (figure 19.1).

Definition: The nitrogen cycle
The nitrogen cycle is a biogeochemical cycle that describes how nitrogen and nitrogen-containing compounds are changed in nature.

Very broadly, the nitrogen cycle is made up of the following processes:

- **Nitrogen fixation** - The process of converting inert nitrogen gas into more useable nitrogen compounds such as ammonia.
- **Nitrification** - The conversion of ammonia into nitrites and then into nitrates, which can be absorbed and used by plants.
- **Denitrification** - The conversion of nitrates back into nitrogen gas in the atmosphere.

We are going to look at each of these processes in more detail.

19.2 Nitrogen fixation

Nitrogen fixation is needed to change gaseous nitrogen into forms such as ammonia that are more useful to living organisms. Some fixation occurs in lightning strikes and in industrial processes,

Figure 19.1: A simplified diagram of the nitrogen cycle

but most fixation is done by different types of bacteria living either in the soil or in parts of the plants.

1. **Biological fixation**

   Some bacteria are able to fix nitrogen. They use an enzyme called *nitrogenase* to combine gaseous nitrogen with hydrogen to form ammonia. The bacteria then use some of this ammonia to produce their own organic compounds, while what is left of the ammonia becomes available in the soil.

   Some of these bacteria are free-living, in other words they live in the soil. Others live in the root nodules of legumes (e.g. soy, peas and beans). Here they form a mutualistic relationship with the plant. The bacteria get carbohydrates (food) from the plant and, in exchange, produce ammonia which can be converted into nitrogen compounds that are essential for the survival of the plant. In nutrient-poor soils, planting lots of legumes can help to enrich the soil with nitrogen compounds.

   A simplified equation for biological nitrogen fixation is:

   $$N_2 + 8H^+ + 8e^- \rightarrow 2NH_3 + H_2$$

   Energy is used in the process, but this is not shown in the above equation.

   Another important source of ammonia in the soil is **decomposition**. When animals and plants die, the nitrogen compounds that were present in them are broken down and converted into ammonia. This process is carried out by decomposition bacteria and fungi in the soil.

2. **Industrial nitrogen fixation**

In the Haber-Bosch process, nitrogen (N_2) is converted together with hydrogen gas (H_2) into ammonia (NH_3) fertiliser. This is an artificial process.

3. **Lightning**

   In the atmosphere, lightning and photons are important in the reaction between nitrogen (N_2) and oxygen (O_2) to form nitric oxide (NO) and then nitrates.

Interesting Fact

It is interesting to note that by cultivating legumes, using the Haber-Bosch process to manufacture chemical fertilisers and increasing pollution from vehicles and industry, humans have more than doubled the amount of nitrogen that would normally be changed from nitrogen gas into a biologically useful form. This has serious environmental consequences.

19.3 Nitrification

Nitrification involves two biological oxidation reactions: firstly, the oxidation of ammonia with oxygen to form nitrite (NO_2^-) and secondly the oxidation of these nitrites into nitrates.

1. $NH_3 + O_2 \rightarrow NO_2^- + 3H^+ + 2e^-$ (production of *nitrites*)
2. $NO_2^- + H_2O \rightarrow NO_3^- + 2H^+ + 2e^-$ (production of *nitrates*)

Nitrification is an important step in the nitrogen cycle in soil because it converts the ammonia (from the nitrogen fixing part of the cycle) into nitrates, which are easily absorbed by the roots of plants. This absorption of nitrates by plants is called **assimilation**. Once the nitrates have been assimilated by the plants, they become part of the plants' proteins. These plant proteins are then available to be eaten by animals. In other words, animals (including humans) obtain their own nitrogen by feeding on plants. Nitrification is performed by bacteria in the soil, called *nitrifying bacteria*.

Activity :: Case Study : Nitrates in drinking water

Read the information below and then carry out your own research to help you answer the questions that follow.

The negatively charged nitrate ion is not held onto soil particles and so can be easily washed out of the soil. This is called **leaching**. In this way, valuable nitrogen can be lost from the soil, reducing the soil's fertility. The nitrates can then accumulate in groundwater, and eventually in drinking water. There are strict regulations that control how much nitrate can be present in drinking water, because nitrates can be reduced to highly reactive nitrites by microorganisms in the gut. Nitrites are absorbed from the gut and bind to haemoglobin (the pigment in blood that helps to transport oxygen around the body). This reduces the ability of the haemoglobin to carry oxygen. In young babies this can lead to respiratory distress, a condition known as "blue baby syndrome".

1. How is nitrate concentration in water measured?
2. What concentration of nitrates in drinking water is considered acceptable? You can use drinking water standards for any part of the world, if you can't find any for South Africa.
3. What is 'blue baby syndrome' and what are the symptoms of the disease?

19.4 Denitrification

Denitrification is the process of reducing nitrate and nitrite into gaseous nitrogen. The process is carried out by *denitrification bacteria*. The nitrogen that is produced is returned to the atmosphere to complete the nitrogen cycle.

The equation for the reaction is:

$$2NO_3^- + 10e^- + 12H^+ \rightarrow N_2 + 6H_2O$$

19.5 Human Influences on the Nitrogen Cycle

Humans have contributed significantly to the nitrogen cycle in a number of ways.

- Both **artificial fertilisation** and the planting of **nitrogen fixing crops**, increase the amount of nitrogen in the soil. In some ways this has positive effects because it increases the fertility of the soil, and means that agricultural productivity is high. On the other hand, however, if there is too much nitrogen in the soil, it can run off into nearby water courses such as rivers, or can become part of the groundwater supply as we mentioned earlier. Increased nitrogen in rivers and dams can lead to a problem called **eutrophication**. Eutrophication is a process where water bodies such as rivers, estuaries, dams and slow-moving streams receive *excess* nutrients (e.g. nitrogen and phosphorus compounds) that stimulate excessive plant growth. Sometimes this can cause certain plant species to be favoured over the others and one species may 'take over' the ecosystem, resulting in a decrease in plant diversity. This is called a 'bloom'. Eutrophication also affects water quality. When the plants die and decompose, large amounts of oxygen are used up and this can cause other animals in the water to die.

Activity :: Case Study : Fertiliser use in South Africa

Refer to the data table below, which shows the average fertiliser use (in kilograms per hectare or kg/ha) over a number of years for South Africa and the world. Then answer the questions that follow:

	1965	1970	1975	1980	1985	1990	1995	2000	2002
SA	27.9	42.2	57.7	80.3	66.6	54.9	48.5	47.1	61.4
World	34.0	48.9	63.9	80.6	86.7	90.9	84.9	88.2	91.9

1. On the same set of axes, draw two line graphs to show how fertiliser use has changed in SA and the world between 1965 and 2002.
2. Describe the trend you see for...
   (a) the world
   (b) South Africa
3. Suggest a reason why the world's fertiliser use has changed in this way over time.
4. Do you see the same pattern for South Africa?
5. Try to suggest a reason for the differences you see in the fertiliser use data for South Africa.
6. One of the problems with increased fertiliser use is that there is a greater chance of nutrient runoff into rivers and dams, and therefore a greater danger of eutrophication. In groups of 5-6, discuss the following questions:
   (a) What could farmers do to try to reduce the risk of nutrient runoff from fields into water systems? Try to think of at least 3 different strategies that they could use.

(b) Imagine you are going to give a presentation on eutrophication to a group of farmers who know nothing about it. How will you educate them about the dangers? How will you convince them that it is in their interests to change their farming practices? Present your ideas to the class.

- **Atmospheric pollution** is another problem. The main culprits are nitrous oxide (N_2O), nitric oxide (NO) and nitrogen dioxide (NO_2). Most of these gases result either from emissions from agricultural soils (and particularly artificial fertilisers), or from the combustion of fossil fuels in industry or motor vehicles. The combustion (burning) of nitrogen-bearing fuels such as coal and oil releases this nitrogen as N_2 or NO gases. Both NO_2 and NO can combine with water droplets in the atmosphere to form **acid rain**. Furthermore, both NO and NO_2 contribute to the depletion of the ozone layer and some are **greenhouse gases**. In high concentrations these gases can contribute towards **global warming**.

19.6 The industrial fixation of nitrogen

A number of industrial processes are able to fix nitrogen into different compounds and then convert these compounds into fertilisers. In the descriptions below, you will see how atmospheric nitrogen is fixed to produce ammonia, how ammonia is then reacted with oxygen to form nitric acid and how nitric acid and ammonia are then used to produce the fertiliser, ammonium nitrate.

- **Preparation of ammonia (NH_3)**

  The industrial preparation of ammonia is known as the **Haber-Bosch process**. At a high pressure and a temperature of approximately 500^0C, and in the presence of a suitable catalyst (usually iron), nitrogen and hydrogen react according to the following equation:

  $$N_2 + 3H_2 \rightarrow 2NH_3$$

  Ammonia is used in the preparation of artficial fertilisers such as $(NH_4)_2SO_4$ and is also used in cleaning agents and cooling installations.

Interesting Fact

Fritz Haber and Carl Bosch were the two men responsible for developing the Haber-Bosch process. In 1918, Haber was awarded the Nobel Prize in Chemistry for his work. The Haber-Bosch process was a milestone in industrial chemistry because it meant that nitrogenous fertilisers were cheaper and much more easily available. At the time, this was very important in providing food for the growing human population.

Haber also played a major role in the development of chemical warfare in World War I. Part of this work included the development of gas masks with absorbent filters. He also led the teams that developed chlorine gas and other deadly gases for use in trench warfare. His wife, Clara Immerwahr, also a chemist, opposed his work on poison gas and committed suicide with his service weapon in their garden. During the 1920s, scientists working at his institute also developed the cyanide gas formulation Zyklon B, which was used as an insecticide and also later, after he left the programme, in the Nazi extermination camps.

Haber was Jewish by birth, but converted from Judaism in order to be more accepted in Germany. Despite this, he was forced to leave the country in 1933 because he was Jewish 'by definition' (his mother was Jewish). He died in 1934 at the age of 65. Many members of his extended family died in the Nazi concentration camps, possibly gassed by Zyklon B.

- **Preparation of nitric acid (HNO_3)**

  Nitric acid is used to prepare fertilisers and explosives. The industrial preparation of nitric acid is known as the **Ostwald process**. The Ostwald process involves the conversion of ammonia into nitric acid in various stages:

  Firstly, ammonia is heated with oxygen in the presence of a platinum catalyst to form nitric oxide and water.

  $$4NH_3(g) + 5O_2(g) \rightarrow 4NO(g) + 6H_2O(g)$$

  Secondly, nitric oxide reacts with oxygen to form nitrogen dioxide. This gas is then readily absorbed by the water to produce nitric acid. A portion of nitrogen dioxide is reduced back to nitric oxide.

  $$2NO(g) + O_2(g) \rightarrow 2NO_2(g)$$
  $$3NO_2(g) + H_2O(l) \rightarrow 2HNO_3(aq) + NO(g)$$

  The NO is recycled, and the acid is concentrated to the required strength by a process called *distillation*.

- **Preparation of ammonium nitrate**

  Ammonium nitrate is used as a fertiliser, as an explosive and also in the preparation of 'laughing gas' which is used as an anaesthetic. Ammonium nitrate is prepared by reacting ammonia with nitric acid:

  $$NH_3 + HNO_3 \rightarrow NH_4NO_3$$

Activity :: Debate : Fertiliser use
Divide the class into two groups to debate the following topic:

Increasing the use of artificial fertilisers is the best solution to meet the growing food needs of the world's human population.

One group should take the position of *agreeing* with the statement, and the other should *disagree*. In your groups, discuss reasons why you have the opinion that you do, and record some notes of your discussion. Your teacher will then explain to you how to proceed with the debate.

19.7 Summary

- **Nitrogen** is essential for life on earth, since it forms part of **amino acids**, **proteins** and **nucleic acids**.
- The **atmosphere** is composed mostly of nitrogen gas, but the gas is **inert**, meaning that it is not available to living organisms in its gaseous form.
- The **nitrogen cycle** describes how nitrogen and nitrogen-containing compounds are changed into different forms in nature.

- The nitrogen cycle consists of three major processes: **nitrogen fixation**, **nitrification** and **denitrification**.
- **Nitrogen fixation** is the conversion of atmospheric nitrogen into compounds such as ammonia, that are more easily used.
- Nitrogen can be fixed **biologically** through the actions of **bacteria**, **industrially** through the **Haber-Bosch process** or by **lightning**.
- **Nitrification** converts ammonia into **nitrites** and **nitrates**, which can be easily **assimilated** by plants.
- **Denitrification** converts nitrites and nitrates back into gaseous nitrogen to complete the nitrogen cycle.
- **Humans** have had a number of **impacts** on the nitrogen cycle. The production of **artificial fertilisers** for example, means that there is a greater chance of runoff into water systems. In some cases, **eutrophication** may occur.
- **Eutrophication** is the enrichment of water systems with excess nutrients, which may stimulate excessive plant growth at the expense of other parts of the ecosystem.
- Many nitrogen gases such as NO, N_2O and NO_2 are released by agricultural soils and artificial fertilisers. These gases may combine with water vapour in the atmosphere and result in **acid rain**. Some of these gases are also greenhouse gases and may contribute towards **global warming**.
- A number of **industrial processes** are used to produce **articifical fertilisers**.
- The **Haber-Bosch process** converts atmsopheric nitrogen into **ammonia**.
- The **Ostwald process** reacts ammonia with oxygen to produce **nitric acid**, which is used in the preparation of fertilisers and explosives.
- If ammonia and nitric acid react, the product is **ammonium nitrate**, which is used as a fertiliser and as an explosive.

Exercise: Summary Exercise

1. Look at the diagram and the descriptions of the nitrogen cycle earlier in the chapter:
   (a) Would you describe the changes that take place in the nitrogen cycle as *chemical* or *physical* changes? Explain your answer.
   (b) Are the changes that take place in the water cycle *physical* or *chemical* changes? Explain your answer.
2. Explain what is meant by each of the following terms:
   (a) nitrogen fixing
   (b) fertiliser
   (c) eutrophication
3. Explain why the fixing of atmospheric nitrogen is so important for the survival of life on earth.
4. Refer to the diagram below and then answer the questions that follow:

(a) Explain the role of *decomposers* in the nitrogen cycle.
(b) If the process taking place at (3) is *nitrification*, then label the processes at (1) and (5).
(c) Identify the nitrogen products at (2) and (4).
(d) On the diagram, indicate the type of *bacteria* that are involved in each stage of the nitrogen cycle.
(e) In industry, what process is used to produce the compound at 2?
(f) Does the diagram above show a 'cycle'? Explain your answer.

5. NO and NO_2 are both nitrogen compounds:
   (a) Explain how each of these compounds is formed?
   (b) What effect does each of these compounds have in the environment?
6. There are a number of arguments both 'for' and 'against' the use of artificial fertilisers. Draw a table to summarise the advantages and disadvantages of their use.

Chapter 20

The Hydrosphere - Grade 10

20.1 Introduction

As far as we know, the Earth we live on is the only planet that is able to support life. Among other things, Earth is just the right distance from the sun to have temperatures that are suitable for life to exist. Also, the Earth's atmosphere has exactly the right type of gases in the right amounts for life to survive. Our planet also has **water** on its surface, which is something very unique. In fact, Earth is often called the 'Blue Planet' because most of it is covered in water. This water is made up of *freshwater* in rivers and lakes, the *saltwater* of the oceans and estuaries, *groundwater* and *water vapour*. Together, all these water bodies are called the **hydrosphere**.

20.2 Interactions of the hydrosphere

It is important to realise that the hydrosphere interacts with other global systems, including the *atmosphere*, *lithosphere* and *biosphere*.

- *Atmosphere*

  When water is heated (e.g. by energy from the sun), it evaporates and forms water vapour. When water vapour cools again, it condenses to form liquid water which eventually returns to the surface by precipitation e.g. rain or snow. This cycle of water moving through the atmosphere, and the energy changes that accompany it, is what drives weather patterns on earth.

- *Lithosphere*

  In the lithosphere (the ocean and continental crust at the Earth's surface), water is an important *weathering* agent, which means that it helps to break rock down into rock fragments and then soil. These fragments may then be transported by water to another place, where they are deposited. This is called *erosion*. These two process i.e. weathering and erosion, help to shape the earth's surface. You can see this for example in rivers. In the upper streams, rocks are eroded and sediments are transported down the river and deposited on the wide flood plains lower down. On a bigger scale, river valleys in mountains have been carved out by the action of water, and cliffs and caves on rocky beach coastlines, are also the result of weathering and erosion by water.

- *Biosphere*

  In the biosphere, land plants absorb water through their roots and then transport this through their vascular (transport) system to stems and leaves. This water is needed in *photosynthesis*, the food production process in plants. Transpiration (evaporation of water from the leaf surface) then returns water back to the atmosphere.

20.3 Exploring the Hydrosphere

The large amount of water on our planet is something quite unique. In fact, about 71% of the earth is covered by water. Of this, almost 97% is found in the oceans as saltwater, about 2.2% occurs as a solid in ice sheets, while the remaining amount (less than 1%) is available as freshwater. So from a human perspective, despite the vast amount of water on the planet, only a very small amount is actually available for human consumption (e.g. drinking water). Before we go on to look more closely at the chemistry of the hydrosphere, we are going to spend some time exploring a part of the hydrosphere, in order to start appreciating what a complex and beautiful part of the world it is.

Activity :: Investigation : Investigating the hydrosphere

1. **Choosing a study site:**
   For this exercise, you can choose any part of the hydrosphere that you would like to explore. This may be a rock pool, a lake, river, wetland or even just a small pond. The guidelines below will apply best to a river investigation, but you can ask similar questions and gather similar data in other areas. When choosing your study site, consider how accessible it is (how easy is it to get to?) and the problems you may experience (e.g. tides, rain).
2. **Collecting data:**
   Your teacher will provide you with the equipment you need to collect the following data. You should have at least one study site where you will collect data, but you might decide to have more if you want to compare your results in different areas. This works best in a river, where you can choose sites down its length.
   (a) *Chemical data*
   Measure and record data such as temperature, pH, conductivity and dissolved oxygen at each of your sites. You may not know exactly what these measurements mean right now, but it will become clearer later in the chapter.
   (b) *Hydrological data*
   Measure the water velocity of the river and observe how the volume of water in the river changes as you move down its length. You can also collect a water sample in a clear bottle, hold it to the light and see whether the water is clear or whether it has particles in it.
   (c) *Biological data*
   What types of animals and plants are found in or near this part of the hydrosphere? Are they specially adapted to their environment?

   Record your data in a table like the one shown below:

	Site 1	**Site 2**	**Site 3**
Temperature			
pH			
Conductivity			
Dissolved oxygen			
Animals and plants			

3. **Interpreting the data:**
   Once you have collected and recorded your data, think about the following questions:
   - How does the data you have collected vary at different sites?
   - Can you explain these differences?
   - What effect do you think *temperature*, *dissolved oxygen* and *pH* have on animals and plants that are living in the hydrosphere?

- Water is seldom 'pure'. It usually has lots of things dissolved (e.g. Mg, Ca and NO_3^- ions) or suspended (e.g. soil particles, debris) in it. Where do these substances come from?
- Are there any human activities near this part of the hydrosphere? What effect could these activities have on the hydrosphere?

20.4 The Importance of the Hydrosphere

It is so easy sometimes to take our hydrosphere for granted, and we seldom take the time to really think about the role that this part of the planet plays in keeping us alive. Below are just some of the very important functions of water in the hydrosphere:

- *Water is a part of living cells*

  Each cell in a living organism is made up of almost 75% water, and this allows the cell to function normally. In fact, most of the chemical reactions that occur in life, involve substances that are dissolved in water. Without water, cells would not be able to carry out their normal functions, and life could not exist.

- *Water provides a habitat*

  The hydrosphere provides an important place for many animals and plants to live. Many gases (e.g. CO_2, O_2), nutrients e.g. nitrate (NO_3^-), nitrite (NO_2^-) and ammonium (NH_4^+) ions, as well as other ions (e.g. Ca^{2+} and Mg^{2+}) are dissolved in water. The presence of these substances is critical for life to exist in water.

- *Regulating climate*

  You may remember from chapter **??** that one of water's unique characteristics is its high *specific heat*. This means that water takes a long time to heat up, and also a long time to cool down. This is important in helping to regulate temperatures on earth so that they stay within a range that is acceptable for life to exist. *Ocean currents* also help to disperse heat.

- *Human needs*

  Humans use water in a number of ways. Drinking water is obviously very important, but water is also used domestically (e.g. washing and cleaning) and in industry. Water can also be used to generate electricity through hydropower.

These are just a few of the very important functions that water plays on our planet. Many of the functions of water relate to its chemistry and to the way in which it is able to dissolve substances in it.

20.5 Ions in aqueous solution

As we mentioned earlier, water is seldom pure. Because of the structure of the water molecule, it is able to dissolve substances in it. This is very important because if water wasn't able to do this, life would not be able to survive. In rivers and the oceans for example, dissolved oxygen means that organisms are still able to respire (breathe). For plants, dissolved nutrients are also available. In the human body, water is able to carry dissolved substances from one part of the body to another.

Many of the substances that dissolve are *ionic*, and when they dissolve they form ions in solution. We are going to look at how water is able to dissolve ionic compounds, and how these ions maintain a balance in the human body, how they affect water hardness, and how specific ions determine the pH of solutions.

20.5.1 Dissociation in water

You may remember from chapter 5 that water is a **polar molecule** (figure 20.1). This means that one part of the molecule has a slightly positive charge and the other part has a slightly negative charge.

Figure 20.1: Water is a polar molecule

It is the polar nature of water that allows ionic compounds to dissolve in it. In the case of sodium chloride (NaCl) for example, the positive sodium ions (Na^+) will be attracted to the negative pole of the water molecule, while the negative chloride ions (Cl^-) will be attracted to the positive pole of the water molecule. In the process, the ionic bonds between the sodium and chloride ions are weakened and the water molecules are able to work their way between the individual ions, surrounding them and slowly dissolving the compound. This process is called **dissociation**. A simplified representation of this is shown in figure 20.2.

Definition: Dissociation
Dissociation in chemistry and biochemistry is a general process in which ionic compounds separate or split into smaller molecules or ions, usually in a reversible manner.

Figure 20.2: Sodium chloride dissolves in water

The dissolution of sodium chloride can be represented by the following equation:

$$NaCl(s) \rightarrow Na^+(aq) + Cl^-(aq)$$

The symbols **s** (solid), **l** (liquid), **g** (gas) and **aq** (material is dissolved in water) are written after the chemical formula to show the state or phase of the material. The dissolution of potassium sulphate into potassium and sulphate ions is shown below as another example:

$$K_2SO_4(s) \rightarrow 2K^+(aq) + SO_4^{2-}(aq)$$

Remember that **molecular** substances (e.g. covalent compounds) may also dissolve, but most will not form ions. One example is sugar.

$$C_6H_{12}O_6(s) \Leftrightarrow C_6H_{12}O_6(aq)$$

There are exceptions to this and some molecular substances *will* form ions when they dissolve. Hydrogen chloride for example can ionise to form hydrogen and chloride ions.

$$HCl(g) \rightarrow H^+(aq) + Cl^-(aq)$$

Interesting Fact

The ability of ionic compounds to dissolve in water is extremely important in the human body! The body is made up of *cells*, each of which is surrounded by a *membrane*. Dissolved ions are found inside and outside of body cells, in different concentrations. Some of these ions are positive (e.g. Mg^{2+}) and some are negative (e.g. Cl^-). If there is a difference in the charge that is inside and outside the cell, then there is a *potential difference* across the cell membrane. This is called the **membrane potential** of the cell. The membrane potential acts like a battery and affects the movement of all charged substances across the membrane. Membrane potentials play a role in muscle functioning, digestion, excretion and in maintaining blood pH, to name just a few. The movement of ions across the membrane can also be converted into an electric signal that can be transferred along *neurons* (nerve cells), which control body processes. If ionic substances were not able to dissociate in water, then none of these processes would be possible! It is also important to realise that our bodies can *lose* ions such as Na^+, K^+, Ca^{2+}, Mg^{2+}, and Cl^-, for example when we sweat during exercise. Sports drinks such as Lucozade and Powerade are designed to replace these lost ions so that the body's normal functioning is not affected.

Exercise: Ions in solution

1. For each of the following, say whether the substance is ionic or molecular.
   (a) potassium nitrate (KNO_3)
   (b) ethanol (C_2H_5OH)
   (c) sucrose sugar ($C_{12}H_{22}O_{11}$
   (d) sodium bromide (NaBr)
2. Write a balanced equation to show how each of the following ionic compounds dissociate in water.
   (a) sodium sulphate (Na_2SO_4)
   (b) potassium bromide (KBr)
   (c) potassium permanganate ($KMNO_4$)
   (d) sodium phosphate (Na_3PO_4)

20.5.2 Ions and water hardness

Definition: Water hardness
Water hardness is a measure of the mineral content of water. Minerals are substances such as calcite, quartz and mica that occur naturally as a result of geological processes.

Hard water is water that has a high mineral content. Water that has a low mineral content is known as **soft water**. If water has a high mineral content, it usually contains high levels of metal ions, mainly calcium (Ca) and magnesium (Mg). The calcium enters the water from either $CaCO_3$ (limestone or chalk) or from mineral deposits of $CaSO_4$. The main source of magnesium is a sedimentary rock called dolomite, $CaMg(CO_3)_2$. Hard water may also contain other metals as well as bicarbonates and sulphates.

Interesting Fact

The simplest way to check whether water is hard or soft is to use the lather/froth test. If the water is very soft, soap will lather more easily when it is rubbed against the skin. With hard water this won't happen. Toothpaste will also not froth well in hard water.

A **water softener** works on the principle of **ion exchange**. Hard water passes through a media bed, usually made of resin beads that are supersaturated with sodium. As the water passes through the beads, the hardness minerals (e.g. calcium and magnesium) attach themselves to the beads. The sodium that was originally on the beads is released into the water. When the resin becomes saturated with calcium and magnesium, it must be recharged. A salt solution is passed through the resin. The sodium replaces the calcium and magnesium, and these ions are released into the waste water and discharged.

20.5.3 The pH scale

The concentration of specific ions in solution, affects whether the solution is acidic or basic. You will learn about acids and bases in chapter 15. Acids and bases can be described as substances that either increase or decrease the concentration of hydrogen (H^+ or H_3O+) ions in a solution. An acid *increases* the hydrogen ion concentration in a solution, while a base *decreases* the hydrogen ion concentration. **pH** is used to measure whether a substance is acidic or basic (alkaline).

Definition: pH
pH is a measure of the acidity or alkalinity of a solution. The pH scale ranges from 0 to 14. Solutions with a pH less than seven are acidic, while those with a pH greater than seven are basic (alkaline). pH 7 is considered neutral.

pH can be calculated using the following equation:

$$pH = -log[H^+]$$

or

$$pH = -log[H_3O^+]$$

The brackets in the above equation are used to show *concentration* in mol.dm^{-3}.

Worked Example 93: pH calculations

Question: Calculate the pH of a solution where the concentration of hydrogen ions is 1×10^{-7} mol.dm^{-3}.

Answer
Step 1 : Determine the concentration of hydrogen ions in mol.dm^{-3}
In this example, the concentration has been given and is 1×10^{-7} mol.dm^{-3}

Step 2 : Substitute this value into the pH equation and calculate the pH value
pH = -log[H^+]
= -log(1×10^{-7})
= 7

Worked Example 94: pH calculations

Question: In a solution of ethanoic acid, the following equilibrium is established:

$$CH_3COOH(aq) + H_2O \Leftrightarrow CH_3COO^-(aq) + H_3O^+$$

The concentration of CH_3COO^- ions is found to be 0.003 mol.dm^{-3}. Calculate the pH of the solution.

Answer
Step 1 : Determine the concentration of hydrogen ions in the solution
According to the balanced equation for this reaction, the mole ratio of CH_3COO^- ions to H_3O^+ ions is the same, therefore the concentration of these two ions in the solution will also be the same. So, [H_3O^+] = 0.003 dm^{-3}.

Step 2 : Substitute this value into the pH equation and calculate the pH value
pH = -log[H_3O^+]
= -log(0.003)
= 2.52

Understanding pH is very important. In living organisms, it is necessary to maintain a constant pH so that chemical reactions can occur under optimal conditions.

Important: It may also be useful for calculations involving the pH scale, to know that the following equation can also be used:

$$[H_3O^+][OH^-] = 1 \times 10^{-14}$$

Interesting Fact

A build up of acid in the human body can be very dangerous. **Lactic acidosis** is a condition caused by the buildup of lactic acid in the body. It leads to acidification of the blood (acidosis) and can make a person very ill. Some of the symptoms of lactic acidosis are deep and rapid breathing, vomiting, and abdominal pain. In the fight against HIV, lactic acidosis is a problem. One of the antiretrovirals (ARV's) that is used in anti-HIV treatment is Stavudine (also known as Zerit or d4T). One of the side effects of Stavudine is lactic acidosis, particularly in overweight women. If it is not treated quickly, it can result in death.

In agriculture, farmers need to know the pH of their soils so that they are able to plant the right kinds of crops. The pH of soils can vary depending on a number of factors such as rainwater, the kinds of rocks and materials from which the soil was formed and also human influences such as pollution and fertilisers. The pH of rain water can also vary and this too has an effect on agriculture, buildings, water courses, animals and plants. Rainwater is naturally acidic because carbon dioxide in the atmosphere combines with water to form carbonic acid. Unpolluted rainwater has a pH of approximately 5.6. However, human activities can alter the acidity of rain and this can cause serious problems such as acid rain.

Exercise: Calculating pH

1. Calculate the pH of each of the following solutions:
   (a) A 0.2 mol.dm^{-3} KOH solution
   (b) A 0.5 mol.dm^{-3} HCl solution
2. What is the concentration (in mol.dm^{-3}) of H_3O^+ ions in a NaOH solution which has a pH of 12?
3. The concentrations of hydronium and hydroxyl ions in a typical sample of seawater are 10^{-8} mol.dm^{-3} and 10^{-6} mol.dm^{-3} respectively.
   (a) Is the seawater acidic or basic?
   (b) What is the pH of the seawater?
   (c) Give a possible explanation for the pH of the seawater.

   (IEB Paper 2, 2002)

20.5.4 Acid rain

The acidity of rainwater comes from the natural presence of three substances (CO_2, NO, and SO_2) in the lowest layer of the atmosphere. These gases are able to dissolve in water and therefore make rain more acidic than it would otherwise be. Of these gases, carbon dioxide (CO_2) has the highest concentration and therefore contributes the most to the natural acidity of rainwater. We will look at each of these gases in turn.

Definition: Acid rain
Acid rain refers to the deposition of acidic components in rain, snow and dew. Acid rain occurs when sulfur dioxide and nitrogen oxides are emitted into the atmosphere, undergo chemical transformations, and are absorbed by water droplets in clouds. The droplets then fall to earth as rain, snow, mist, dry dust, hail, or sleet. This increases the acidity of the soil, and affects the chemical balance of lakes and streams.

1. **Carbon dioxide**

   Carbon dioxide reacts with water in the atmosphere to form **carbonic acid** (H_2CO_3).

   $$CO_2 + H_2O \rightarrow H_2CO_3$$

   The carbonic acid dissociates to form hydrogen and hydrogen carbonate ions. It is the presence of hydrogen ions that lowers the pH of the solution, making the rain acidic.

   $$H_2CO_3 \rightarrow H^+ + HCO_3^-$$

2. **Nitric oxide**

   Nitric oxide (NO) also contributes to the natural acidity of rainwater and is formed during lightning storms when nitrogen and oxygen react. In air, NO is oxidised to form nitrogen dioxide (NO_2). It is the nitrogen dioxide which then reacts with water in the atmosphere to form **nitric acid** (HNO_3).

   $$3NO_2(g) + H_2O \rightarrow 2HNO_3(aq) + NO(g)$$

   The nitric acid dissociates in water to produce hydrogen ions and nitrate ions. This again lowers the pH of the solution, making it acidic.

   $$HNO_3 \rightarrow H^+ + NO_3^-$$

3. **Sulfur dioxide**

   Sulfur dioxide in the atmosphere first reacts with oxygen to form sulfur trioxide, before reacting with water to form **sulfuric acid**.

   $$2SO_2 + O_2 \rightarrow 2SO_3$$

   $$SO_3 + H_2O \rightarrow H_2SO_4$$

   Sulfuric acid dissociates in a similar way to the previous reactions.

   $$H_2SO_4 \rightarrow HSO_4^- + H^+$$

Although these reactions do take place naturally, human activities can greatly increase the concentration of these gases in the atmosphere, so that rain becomes far more acidic than it would otherwise be. The burning of fossil fuels in industries, vehicles etc is one of the biggest culprits. If the acidity of the rain drops below 5, it is referred to as **acid rain**.

Acid rain can have a very damaging effect on the environment. In rivers, dams and lakes, increased acidity can mean that some species of animals and plants will not survive. Acid rain can also degrade soil minerals, producing metal ions that are washed into water systems. Some of these ions may be toxic e.g. Al^{3+}. From an economic perspective, altered soil pH can drastically affect agricultural productivity.

Acid rain can also affect buildings and monuments, many of which are made from marble and limestone. A chemical reaction takes place between $CaCO_3$ (limestone) and sulfuric acid to produce aqueous ions which can be easily washed away. The same reaction can occur in the lithosphere where limestone rocks are present e.g. limestone caves can be eroded by acidic rainwater.

$$H_2SO_4 + CaCO_3 \rightarrow CaSO_4.H_2O + CO_2$$

Activity :: Investigation : Acid rain
You are going to test the effect of 'acid rain' on a number of substances.
Materials needed:
samples of chalk, marble, zinc, iron, lead, dilute sulfuric acid, test tubes, beaker, glass dropper
Method:

1. Place a small sample of each of the following substances in a separate test tube: chalk, marble, zinc, iron and lead
2. To each test tube, add a few drops of dilute sulfuric acid.
3. Observe what happens and record your results.

Discussion questions:

- In which of the test tubes did reactions take place? What happened to the sample substances?
- What do your results tell you about the effect that acid rain could have on each of the following: buildings, soils, rocks and geology, water ecosystems?
- What precautions could be taken to reduce the potential impact of acid rain?

20.6 Electrolytes, ionisation and conductivity

Conductivity in aqueous solutions, is a measure of the ability of water to conduct an electric current. The more **ions** there are in the solution, the higher its conductivity.

Definition: Conductivity
Conductivity is a measure of a solution's ability to conduct an electric current.

20.6.1 Electrolytes

An **electrolyte** is a material that *increases* the conductivity of water when dissolved in it. Electrolytes can be further divided into **strong electrolytes** and **weak electrolytes**.

Definition: Electrolyte
An electrolyte is a substance that contains free ions and behaves as an electrically conductive medium. Because they generally consist of ions in solution, electrolytes are also known as ionic solutions.

1. **Strong electrolytes**

   A strong electrolyte is a material that ionises completely when it is dissolved in water:

   $$AB(s,l,g) \rightarrow A^{+}(aq) + B^{-}(aq)$$

   This is a **chemical change** because the original compound has been split into its component ions and bonds have been broken. In a strong electrolyte, we say that the *extent of ionisation* is high. In other words, the original material dissociates completely so that there is a high concentration of ions in the solution. An example is a solution of potassium nitrate:

   $$KNO_3(s) \rightarrow K^{+}(aq) + NO_3^{-}(aq)$$

2. **Weak electrolytes**

   A weak electrolyte is a material that goes into solution and will be surrounded by water molecules when it is added to water. However, not *all* of the molecules will dissociate into ions. The *extent of ionisation* of a weak electrolyte is low and therefore the concentration of ions in the solution is also low.

$$AB(s,l,g) \rightarrow AB(aq) \Leftrightarrow A^{+}(aq) + B^{-}(aq)$$

   The following example shows that, in the final solution of a weak electrolyte, some of the original compound *plus* some dissolved ions are present.

$$C_2H_3O_2H(l) \rightarrow C_2H_3O_2H \Leftrightarrow C_2H_3O_2^{-}(aq) + H^{+}(aq)$$

20.6.2 Non-electrolytes

A **non-electrolyte** is a material that does not increase the conductivity of water when dissolved in it. The substance goes into solution and becomes surrounded by water molecules, so that the molecules of the chemical become separated from each other. However, although the substance does dissolve, it is not changed in any way and no chemical bonds are broken. The change is a **physical change**. In the oxygen example below, the reaction is shown to be reversible because oxygen is only partially soluble in water and comes out of solution very easily.

$$C_2H_5OH(l) \rightarrow C_2H_5OH(aq)$$

$$O_2(g) \Leftrightarrow O_2(aq)$$

20.6.3 Factors that affect the conductivity of water

The conductivity of water is therefore affected by the following factors:

- The **type of substance** that dissolves in water

  Whether a material is a strong electrolyte (e.g. potassium nitrate, KNO_3), a weak electrolyte (e.g. acetate, $C_2H_3O_2H$) or a non-electrolyte (e.g. sugar, alcohol, oil) will affect the conductivity of water because the concentration of ions in solution will be different in each case.

- The **concentration of ions** in solution

  The higher the concentration of ions in solution, the higher its conductivity will be.

- **Temperature**

  The warmer the solution the higher the solubility of the material being dissolved, and therefore the higher the conductivity as well.

Activity :: Experiment : Electrical conductivity

Aim:

To investigate the electrical conductivities of different substances and solutions.

Apparatus:

solid salt (NaCl) crystals; different liquids such as distilled water, tap water, seawater, benzene and alcohol; solutions of salts e.g. NaCl, KBr; a solution of an acid (e.g. HCl) and a solution of a base (e.g. NaOH); torch cells; ammeter; conducting wire, crocodile clips and 2 carbon rods.

Method:

Set up the experiment by connecting the circuit as shown in the diagram below. In the diagram, 'X' represents the substance or solution that you will be testing. When you are using the solid crystals, the crocodile clips can be attached directly to each end of the crystal. When you are using solutions, two carbon rods are placed into the liquid, and the clips are attached to each of the rods. In each case, complete the circuit and allow the current to flow for about 30 seconds. Observe whether the ammeter shows a reading.

Results:

Record your observations in a table similar to the one below:

Test substance	Ammeter reading

What do you notice? Can you explain these observations?

Remember that for electricity to flow, there needs to be a movement of charged particles e.g. ions. With the solid NaCl crystals, there was no flow of electricity recorded on the ammeter. Although the solid is made up of ions, they are held together very tightly within the crystal lattice, and therefore no current will flow. Distilled water, benzene and alcohol also don't conduct a current because they are *covalent compounds* and therefore do not contain ions.

The ammeter should have recorded a current when the salt solutions and the acid and base solutions were connected in the circuit. In solution, salts *dissociate* into their ions, so that these are free to move in the solution. Acids and bases behave in a similar way, and dissociate to form hydronium and oxonium ions. Look at the following examples:

$$\mathrm{KBr} \rightarrow \mathrm{K^+} + \mathrm{Br^-}$$
$$\mathrm{NaCl} \rightarrow \mathrm{Na^+} + \mathrm{Cl^-}$$
$$\mathrm{HCl} + \mathrm{H_2O} \rightarrow \mathrm{H_3O^+} + \mathrm{Cl^-}$$
$$\mathrm{NaOH} \rightarrow \mathrm{Na^+} + \mathrm{OH^-}$$

Conclusions:

Solutions that contain free-moving ions are able to conduct electricity because of the movement of charged particles. Solutions that do not contain free-moving ions do not conduct electricity.

Interesting Fact

Conductivity in streams and rivers is affected by the geology of the area where the water is flowing through. Streams that run through areas with granite bedrock tend to have lower conductivity because granite is made of materials that do not ionise when washed into the water. On the other hand, streams that run through areas with clay soils tend to have higher conductivity because the materials ionise when they are washed into the water. Pollution can also affect conductivity. A failing sewage system or an inflow of fertiliser runoff would raise the conductivity because of the presence of chloride, phosphate, and nitrate (ions) while an oil spill (non-ionic) would lower the conductivity. It is very important that conductivity is kept within a certain acceptable range so that the organisms living in these water systems are able to survive.

20.7 Precipitation reactions

Sometimes, ions in solution may react with each other to form a new substance that is *insoluble*. This is called a **precipitate**.

Definition: Precipitate
A precipitate is the solid that forms in a solution during a chemical reaction.

Activity :: Demonstration : The reaction of ions in solution
Apparatus and materials:
4 test tubes; copper(II) chloride solution; sodium carbonate solution; sodium sulphate solution

Method:

1. Prepare 2 test tubes with approximately 5 ml of dilute Cu(II)chloride solution in each
2. Prepare 1 test tube with 5 ml sodium carbonate solution
3. Prepare 1 test tube with 5 ml sodium sulphate solution
4. Carefully pour the sodium carbonate solution into one of the test tubes containing copper(II) chloride and observe what happens
5. Carefully pour the sodium sulphate solution into the second test tube containing copper(II) chloride and observe what happens

Results:

1. A light blue precipitate forms when sodium carbonate reacts with copper(II) chloride
2. No precipitate forms when sodium sulphate reacts with copper(II) chloride

---

It is important to understand what happened in the previous demonstration. We will look at what happens in each reaction, step by step.

1. **Reaction 1:** Sodium carbonate reacts with copper(II) chloride

   When these compounds react, a number of ions are present in solution: Cu^{2+}, Cl^{-}, Na^{+} and CO_3^{2-}.

   Because there are lots of ions in solution, they will collide with each other and may recombine in different ways. The product that forms may be insoluble, in which case a precipitate will form, or the product will be soluble, in which case the ions will go back into solution. Let's see how the ions in this example could have combined with each other:

   $\mathrm{Cu^{2+} + CO_3^{2-} \rightarrow CuCO_3}$
   $\mathrm{Cu^{2+} + 2Cl^{-} \rightarrow CuCl_2}$
   $\mathrm{Na^{+} + Cl^{-} \rightarrow NaCl}$
   $\mathrm{Na^{+} + CO_3^{2-} \rightarrow Na_2CO_3}$

   You can automatically exclude the reactions where sodium carbonate and copper(II) chloride are the products because these were the initial reactants. You also know that sodium chloride (NaCl) is soluble in water, so the remaining product (copper carbonate) must be the one that is insoluble. It is also possible to look up which salts are soluble and which are insoluble. If you do this, you will find that most carbonates are insoluble, therefore the precipitate that forms in this reaction must be $CuCO_3$. The reaction that has taken place between the ions in solution is as follows:

   $$2Na^{+} + CO_3^{2-} + Cu^{2+} + 2Cl^{-} \rightarrow CuCO_3 + 2Na^{+} + 2Cl^{-}$$

2. **Reaction 2:** Sodium sulphate reacts with copper(II) chloride

   The ions that are present in solution are Cu^{2+}, Cl^{-}, Na^{+} and SO_4^{2-}.

   The ions collide with each other and may recombine in different ways. The possible combinations of the ions are as follows:

   $\mathrm{Cu^{2+} + SO_4^{2-} \rightarrow CuSO_4}$
   $\mathrm{Cu^{2+} + 2Cl^{-} \rightarrow CuCl_2}$
   $\mathrm{Na^{+} + Cl^{-} \rightarrow NaCl}$
   $\mathrm{Na^{+} + SO_4^{2-} \rightarrow Na_2SO_4}$

   If we look up which of these salts are soluble and which are insoluble, we see that most chlorides and most sulphates are soluble. This is why no precipitate forms in this second reaction. Even when the ions recombine, they immediately separate and go back into solution. The reaction that has taken place between the ions in solution is as follows:

   $$2Na^{+} + SO_4^{2-} + Cu^{2+} + 2Cl^{-} \rightarrow 2Na^{+} + SO_4^{2-} + Cu^{2+} + 2Cl^{-}$$

Table 20.1 shows some of the general rules about the solubility of different salts based on a number of investigations:

Table 20.1: General rules for the solubility of salts

Salt	Solubility
Nitrates	All are soluble
Potassium, sodium and ammonium salts	All are soluble
Chlorides	All are soluble except silver chloride, lead(II)chloride and mercury(II)chloride
Sulphates	All are soluble except lead(II)sulphate, barium sulphate and calcium sulphate
Carbonates	All are insoluble except those of potassium, sodium and ammonium

20.8 Testing for common anions in solution

It is also possible to carry out tests to determine which ions are present in a solution.

20.8.1 Test for a chloride

Prepare a solution of the unknown salt using distilled water and add a small amount of **silver nitrate** solution. If a white precipitate forms, the salt is either a chloride or a carbonate.

$$\mathrm{Cl^- + Ag^+ + NO_3^- \rightarrow AgCl + NO_3^-} \text{ (AgCl is white precipitate)}$$
$$\mathrm{CO_3^{2-} + 2Ag^+ + 2NO_3^- \rightarrow Ag_2CO_3 + 2NO_2} \text{ (Ag_2CO_3 is white precipitate)}$$

The next step is to treat the precipitate with a small amount of **concentrated nitric acid**. If the precipitate remains unchanged, then the salt is a chloride. If carbon dioxide is formed, and the precipitate disappears, the salt is a carbonate.

$$\mathrm{AgCl + HNO_3 \rightarrow} \text{ (no reaction; precipitate is unchanged)}$$
$$\mathrm{Ag_2CO_3 + 2HNO_3 \rightarrow 2AgNO_3 + H_2O + CO_2} \text{ (precipitate disappears)}$$

20.8.2 Test for a sulphate

Add a small amount of barium chloride solution to a solution of the test salt. If a white precipitate forms, the salt is either a sulphate or a carbonate.

$$\mathrm{SO_4^{2-} + Ba^{2+} + Cl^- \rightarrow BaSO_4 + Cl^-} \text{ ($BaSO_4$ is a white precipitate)}$$
$$\mathrm{CO_3^{2-} + Ba^{2+} + Cl^- \rightarrow BaCO_3 + Cl^-} \text{ ($BaCO_3$ is a white precipitate)}$$

If the precipitate is treated with nitric acid, it is possible to distinguish whether the salt is a sulphate or a carbonate (as in the test for a chloride).

$$\mathrm{BaSO_4 + HNO_3 \rightarrow} \text{ (no reaction; precipitate is unchanged)}$$
$$\mathrm{BaCO_3 + 2HNO_3 \rightarrow Ba(NO_3)_2 + H_2O + CO_2} \text{ (precipitate disappears)}$$

20.8.3 Test for a carbonate

If a sample of the dry salt is treated with a small amount of acid, the production of carbon dioxide is a positive test for a carbonate.

$$\text{Acid} + CO_3^{2-} \rightarrow CO_2$$

If the gas is passed through limewater and the solution becomes milky, the gas is carbon dioxide.

$$Ca(OH)_2 + CO_2 \rightarrow CaCO_3 + H_2O \text{ (It is the insoluble CaCO}_3\text{ precipitate that makes the limewater go milky)}$$

20.8.4 Test for bromides and iodides

As was the case with the chlorides, the bromides and iodides also form precipitates when they are reacted with silver nitrate. Silver chloride is a white precipitate, but the silver bromide and silver iodide precipitates are both pale yellow. To determine whether the precipitate is a bromide or an iodide, we use chlorine water and carbon tetrachloride (CCl_4).

Chlorine water frees bromine gas from the bromide, and colours the carbon tetrachloride a reddish brown.

Chlorine water frees iodine gas from an iodide, and colours the carbon tetrachloride is coloured purple.

Exercise: Precipitation reactions and ions in solution

1. Silver nitrate ($AgNO_3$) reacts with potassium chloride (KCl) and a white precipitate is formed.
   (a) Write a balanced equation for the reaction that takes place.
   (b) What is the name of the insoluble salt that forms?
   (c) Which of the salts in this reaction are soluble?
2. Barium chloride reacts with sulfuric acid to produce barium sulphate and hydrochloric acid.
   (a) Write a balanced equation for the reaction that takes place.
   (b) Does a precipitate form during the reaction?
   (c) Describe a test that could be used to test for the presence of barium sulphate in the products.
3. A test tube contains a clear, colourless salt solution. A few drops of silver nitrate solution are added to the solution and a pale yellow precipitate forms. Which one of the following salts was dissolved in the original solution?
   (a) NaI
   (b) KCl
   (c) K_2CO_3
   (d) Na_2SO_4

   (IEB Paper 2, 2005)

20.9 Threats to the Hydrosphere

It should be clear by now that the hydrosphere plays an extremely important role in the survival of life on Earth, and that the unique properties of water allow various important chemical processes to take place which would otherwise not be possible. Unfortunately for us however, there are a number of factors that threaten our hydrosphere, and most of these threats are because of human activities. We are going to focus on two of these issues: **overuse** and **pollution** and look at ways in which these problems can possibly be overcome.

1. **Overuse of water**

   We mentioned earlier that only a very small percentage of the hydrosphere's water is available as freshwater. However, despite this, humans continue to use more and more water to the point where water *consumption* is fast approaching the amount of water that is *available*. The situation is a serious one, particularly in countries such as South Africa which are naturally dry, and where water resources are limited. It is estimated that between 2020 and 2040, water supplies in South Africa will no longer be able to meet the growing demand for water in this country. This is partly due to population growth, but also because of the increasing needs of industries as they expand and develop. For each of us, this should be a very scary thought. Try to imagine a day without water...difficult isn't it? Water is so much a part of our lives, that we are hardly aware of the huge part that it plays in our daily lives.

Activity :: Discussion : Creative water conservation

As populations grow, so do the demands that are placed on dwindling water resources. While many people argue that building dams helps to solve this water-shortage problem, the reality is that dams are only a temporary solution, and that they often end up doing far more ecological damage than good. The only sustainable solution is to reduce the *demand* for water, so that water supplies are sufficient to meet this. The more important question then is how to do this.

Discussion:

Divide the class into groups, so that there are about five people in each. Each group is going to represent a different sector within society. Your teacher will tell you which sector you belong to from the following: Farming, industry, city management or civil society (i.e. you will represent the ordinary 'man on the street'). In your groups, discuss the following questions as they relate to the group of people you represent: (Remember to take notes during your discussions, and nominate a spokesperson to give feedback to the rest of the class on behalf of your group)

- What steps could be taken by your group to conserve water?
- Why do you think these steps are *not* being taken?
- What incentives do you think could be introduced to encourage this group to conserve water more efficiently?

---

2. **Pollution**

   Pollution of the hydrosphere is also a major problem. When we think of pollution, we sometimes only think of things like plastic, bottles, oil and so on. But any chemical that is present in the hydrosphere in an amount that is not what it should be is a pollutant. Animals and plants that live in the hydrosphere are specially adapted to surviving within a certain range of conditions. If these conditions are changed (e.g. through pollution), these organisms may not be able to survive. Pollution then, can affect entire aquatic ecosystems. The most common forms of pollution in the hydrosphere are *waste products* from humans and from industries, *nutrient pollution* e.g. fertiliser runoff which causes eutrophication (this will be discussed in a later section) and toxic trace elements such as aluminium, mercury and copper to name a few. Most of these elements come from mines or from industries.

It is important to realise that our hydrosphere exists in a delicate balance with other systems, and that disturbing this balance can have serious consequences for life on this planet.

Activity :: Group Project : School Action Project

There is a lot that can be done within a school to save water. As a class, discuss what actions could be taken by your class to make people more aware of how important it is to conserve water.

20.10 Summary

- The **hydrosphere** includes all the water that is on Earth. Sources of water include freshwater (e.g. rivers, lakes), saltwater (e.g. oceans), groundwater (e.g. boreholes) and water vapour. Ice (e.g. glaciers) is also part of the hydrosphere.
- The hydrosphere interacts with other **global systems**, including the atmosphere, lithosphere and biosphere.
- The hydrosphere has a number of important **functions**. Water is a part of all living cells, it provides a habitat for many living organisms, it helps to regulate climate, and it is used by humans for domestic, industrial and other use.
- The **polar** nature of water means that **ionic compounds** dissociate easily in aqueous solution into their component ions.
- **Ions** in solution play a number of roles. In the human body for example, ions help to regulate the internal environment (e.g. controlling muscle function, regulating blood pH). Ions in solution also determine water hardness and pH.
- **Water hardness** is a measure of the mineral content of water. Hard water has a high mineral concentration and generally also a high concentration of metal ions e.g. calcium and magnesium. The opposite is true for soft water.
- **pH** is a measure of the concentration of hydrogen ions in solution. The formula used to calculate pH is as follows:

  $$pH = -\log[H_3O^+] \text{ or } pH = -\log[H^+]$$

  A solution with a pH less than 7 is considered acidic and more than 7 is considered basic (or alkaline). A neutral solution has a pH of 7.
- Gases such as CO_2, NO_2 and SO^2 dissolve in water to form weak acid solutions. Rain is naturally acidic because of the high concentrations of carbon dioxide in the atmosphere. Human activities such as burning fossil fuels, increase the concentration of these gases in the atmosphere, resulting in **acid rain**.
- **Conductivity** is a measure of a solution's ability to conduct an electric current.
- An **electrolyte** is a substance that contains free ions, and is therefore able to conduct an electric current. Electrolytes can be divided into **strong** and **weak** electrolytes, based on the extent to which the substance ionises in solution.
- A **non-electrolyte** cannot conduct an electric current because it dooes not contain free ions.
- The **type of substance**, the **concentration of ions** and the **temperature** of the solution, affect its conductivity.

- A **precipitate** is formed when ions in solution react with each other to form an insoluble product. Solubility 'rules' help to identify the precipitate that has been formed.
- A number of tests can be used to identify whether certain **anions** are present in a solution.
- Despite the importance of the hydrosphere, a number of factors threaten it. These include **overuse** of water, and **pollution**.

Exercise: Summary Exercise

1. Give one word for each of the following descriptions:
   (a) the change in phase of water from a gas to a liquid
   (b) a charged atom
   (c) a term used to describe the mineral content of water
   (d) a gas that forms sulfuric acid when it reacts with water
2. Match the information in column A with the information in column B by writing only the letter (A to I) next to the question number (1 to 7)

Column A	Column B
1. A polar molecule	A. H_2SO_4
2. molecular solution	B. $CaCO_3$
3. Mineral that increases water hardness	C. NaOH
4. Substance that increases the hydrogen ion concentration	D. salt water
5. A strong electrolyte	E. calcium
6. A white precipitate	F. carbon dioxide
7. A non-conductor of electricity	G. potassium nitrate
	H. sugar water
	I. O_2

3. For each of the following questions, choose the one correct answer from the list provided.
   (a) Which one of the following substances does not conduct electricity in the solid phase but is an electrical conductor when molten?
      i. Cu
      ii. $PbBr_2$
      iii. H_2O
      iv. I_2
      (IEB Paper 2, 2003)
   (b) The following substances are dissolved in water. Which one of the solutions is basic?
      i. sodium nitrate
      ii. calcium sulphate
      iii. ammonium chloride
      iv. potassium carbonate
      (IEB Paper 2, 2005)
4. The concentration of hydronium and hydroxyl ions in a typical sample of seawater are 10^{-8} and 10^{-6} respectively.
   (a) Is the seawater acidic or basic?
   (b) Calculate the pH of this seawater.
5. Three test tubes (X, Y and Z) each contain a solution of an unknown potassium salt. The following observations were made during a practical investigation to identify the solutions in the test tubes:
   A: A white precipitate formed when silver nitrate ($AgNO_3$) was added to test tube Z.

B: A white precipitate formed in test tubes X and Y when barium chloride ($BaCl_2$) was added.

C: The precipitate in test tube X dissolved in hydrochloric acid (HCl) and a gas was released.

D: The precipitate in test tube Y was insoluble in hydrochloric acid.

(a) Use the above information to identify the solutions in each of the test tubes X, Y and Z.

(b) Write a chemical equation for the reaction that took place in test tube X before hydrochloric acid was added.

(DoE Exemplar Paper 2 2007)

Chapter 21

The Lithosphere - Grade 11

(NOTE TO SELF: Need map showing major mining areas in SA)

21.1 Introduction

If we were to cut the Earth in half we would see that our planet is made up of a number of layers, namely the **core** at the centre, the **mantle** and the outer **crust** (figure 21.1). The core is made up mostly of iron. The mantle, which lies between the core and the crust, consists of molten rock, called **magma** which moves continuously because of convection currents. The crust is the thin, hard outer layer that 'floats' on the magma of the mantle. It is the upper part of the mantle and the crust that make up the **lithosphere** ('lith' means 'types of stone' and 'sphere' refers to the round shape of the earth). Together, the lithosphere, hydrosphere and atmosphere make up the world as we know it.

Figure 21.1: A cross-section through the earth to show its different layers

Definition: Lithosphere
The lithosphere is the solid outermost shell of our planet. The lithosphere includes the crust and the upper part of the mantle, and is made up of material from both the continents and the oceans on the Earth's surface.

Earlier chapters have focused on the hydrosphere (chapter 20) and the atmosphere (chapter 22). The lithosphere is also very important, not only because it is the surface on which we live, but also because humans gain many valuable resources from this part of the planet.

21.2 The chemistry of the earth's crust

The crust is made up of about 80 elements, which occur in over 2000 different compounds and minerals. However, most of the mass of crustal material is made up of only 8 of these elements. These are oxygen (O), silica (Si), aluminium (Al), iron (Fe), calcium (Ca), sodium (Na), potassium (K) and magnesium (Mg). These metal elements are seldom found in their pure form, but are usually part of other more complex **minerals**. A mineral is a compound that is formed through geological processes, which give it a particular structure. A mineral could be a pure element, but more often minerals are made up of many different elements combined. *Quartz* is just one example. It is a mineral that is made up of silica and oxygen. Some more examples are shown in table 21.1.

Definition: Mineral
Minerals are natural compounds formed through geological processes. The term 'mineral' includes both the material's chemical composition and its structure. Minerals range in composition from pure elements to complex compounds.

Table 21.1: Table showing examples of minerals and their chemistry

Mineral	Chemistry	Comments
Quartz	SiO_2 (silicon dioxide)	Quartz is used for glass, in electrical components, optical lenses and in building stone
Gold	Au (pure element) or $AuTe_2$ (Calaverite, a gold mineral)	Gold is often found in a group of minerals called the *tellurides*. Calaverite is a mineral that belongs to this group, and is the most common gold-bearing mineral. Gold has an affinity for tellurium (Te).
Hematite	Fe_2O_3 (iron oxide)	Iron usually occurs in iron oxide minerals or as an alloy of iron and nickel.
Orthoclase	$KAlSi_3O_8$ (potassium aluminium silicate)	Orthoclase belongs to the *feldspar* group of minerals.
Copper	Cu (pure element) or $Cu_2(CO_3)(OH)_2$ (malachite or copper carbonate hydroxide)	copper can be mined as a pure element or as a mineral such as malachite.

A **rock** is a combination of one or more minerals. *Granite* for example, is a rock that is made up of minerals such as SiO_2, Al_2O_3, CaO, K_2O, Na_2O and others. There are three different types of rocks, **igneous**, **sedimentary** and **metamorphic**. Igneous rocks (e.g. granite, basalt) are formed when magma is brought to the earth's surface as lava, and then solidifies. Sedimentary rocks (e.g. sandstone, limestone) form when rock fragments, organic matter or other sediment particles are deposited and then compacted over time until they solidify. Metamorphic rock is formed when any other rock types are subjected to intense heat and pressure over a period of time. Examples include slate and marble.

Many of the elements that are of interest to us (e.g. gold, iron, copper), are unevenly distributed in the lithosphere. In places where these elements are abundant, it is profitable to extract them (e.g. through mining) for economic purposes. If their concentration is very low, then the cost of extraction becomes more than the money that would be made if they were sold. Rocks that contain valuable minerals are called **ores**. As humans, we are particularly interested in the ores that contain metal elements, and also in those minerals that can be used to produce energy.

Definition: Ore
An ore is a volume of rock that contains minerals which make it valuable for mining.

Interesting Fact

A **gemstone** (also sometimes called a **gem** or **semi-precious stone**), is a highly attractive and valuable piece of mineral which, when cut and polished, is used in jewelry and other adornments. Examples of gemstones are amethyst, diamond, cat's eye and sapphire.

Exercise: Rocks and minerals

1. Where are most of the earth's minerals concentrated?
2. Explain the difference between a *mineral*, a *rock* and an *ore*.
3. Carry out your own research to find out which elements are found in the following minerals:
   (a) gypsum
   (b) calcite
   (c) topaz
4. Which minerals are found in the following rocks?
   (a) basalt
   (b) sandstone
   (c) marble

21.3 A brief history of mineral use

Many of the minerals that are important to humans are **metals** such as gold, aluminium, copper and iron. Throughout history, metals have played a very important role in making jewelery, tools, weapons and later machinery and other forms of technology. We have become so used to having metals around us that we seldom stop to think what life might have been like before metals were discovered. During the **Stone Age** for example, **stones** were used to make tools. Slivers of stone were cut from a rock and then sharpened. In Africa, some of the stone tools that have been found date back to 2.5 million years ago!

It was the discovery of *metals* that led to some huge advances in agriculture, warfare, transport and even cookery. One of the first metals to be discovered was **gold**, probably because of its beautiful shiny appearance. Gold was used mostly to make jewelery because it was too soft to make harder tools. Later, **copper** became an important metal because it could be hammered into shape, and it also lasted a lot longer than the stone that had previously been used in knives, cooking utensils and weapons. Copper can also be melted and then put into a mould to re-shape it. This is known as **casting**.

At about the time that copper was in widespread use, it was discovered that if certain kinds of stones are heated to high enough temperatures, liquid metals flow from them. These rocks are

ores, and contain the metal minerals inside them. The process of heating mineral ores in this way is called **smelting**. It was also found that ores do not only occur at the earth's surface, but also deep *below* it. This discovery led to the beginning of **mining**.

But humans' explorations into the world of metals did not end here! In some areas, the ores of *iron* and *tin* were found close together. The cast alloy of these two metals is **bronze**. Bronze is a very useful metal because it produces a sharper edge than copper. Another important discovery was that of **iron**. Iron is the most abundant metal at the earth's surface but it is more difficult to work with than copper or tin. It is very difficult to extract iron from its ore because it has an extremely high melting point, and only specially designed furnaces are able to produce the temperatures that are needed. An important discovery was that if iron is heated in a furnace with *charcoal*, some of the carbon in the charcoal is transferred to the iron, making the metal even harder. If this hot metal has its temperature reduced very suddenly, it becomes even harder and produces **steel**. Today, steel is a very important part of industry and construction.

Interesting Fact

Originally it was believed that much of Africa's knowledge of metals and their uses was from the Middle East. But this may not be the case. More recent studies have shown that iron was used far earlier than it would have been if knowledge of this metal had started in the Middle East. Many metal technologies may in fact have developed independently in Africa and in many African countries, metals have an extremely important place in society. In Nigeria's Yoruba country for example, iron has divine status because it is used to make instruments for survival. 'Ogun', the God of Iron, is seen as the protector of the kingdom.

21.4 Energy resources and their uses

Apart from minerals and ores, the products of the lithosphere are also important in meeting our energy needs.

Coal is one of the most important fuels that is used in the production of electricity. Coal is formed from organic material when plants and animals decompose, leaving behind organic remains that accumulate and become compacted over millions of years under sedimentary rock. The layers of compact organic material that can be found between sedimentary layers, are coal. When coal is burned, a large amount of heat energy is released, which is used to produce electricity. South Africa is the world's sixth largest coal producer, with Mpumalanga contributing about 83

Another element that is found in the crust, and which helps to meet our energy needs, is **uranium**. Uranium produces energy through the process of *nuclear fission*. Neutrons are aimed at the nucleii of the uranium atoms in order to split them. When the nucleus of a uranium atom is split, a large amount of energy is released as heat. This heat is used to produce steam, which turns turbines to generate electricity. Uranium is produced as a by-product of gold in some mines in the Witwatersrand, and as a by-product in some copper mines, for example in Phalaborwa. This type of nuclear power is relatively environmentally friendly since it produces low gas emissions. However, the process does produce small amounts of radioactive wastes , which must be carefully disposed of in order to prevent contamination.

Oil is another product of the lithosphere which is critical in meeting our fuel needs. While most of South Africa's oil is imported and then processed at a refinery in either Durban, Cape Town or Sasolburg, some is extracted from coal. The technology behind this type of extraction has largely been developed by SASOL (South African Coal, Oil and Gas Corporation). Oil, like coal,

is organic in origin and normally forms from organic deposits on the ocean floor. Oil requires unique geological and geochemical conditions in order to be produced. Part of this process involves the burial of organic-rich sediments under extremely high temperatures and pressures. The oil that is produced is then pushed out into nearby sedimentary layers. Oil will then move upwards until it is trapped by an impermeable rock layer. It accumulates here, and can then be accessed by oil rigs and other advanced equipment.

21.5 Mining and Mineral Processing: Gold

21.5.1 Introduction

Gold was discovered in South Africa in the late 1800's and since then has played a very important role in South Africa's *history* and *economy*. Its discovery brought many foreigners into South Africa, who were lured by the promises of wealth. They set up small mining villages, which later grew into larger settlements, towns and cities. One of the first of these settlements was the beginning of present-day Johannesburg, also known as 'Egoli' or 'Place of Gold'.

Most of South Africa's gold is concentrated in the 'Golden Arc', which stretches from Johannesburg to Welkom. Geologists believe that, millions of years ago, this area was a massive inland lake. Rivers feeding into this lake brought sand, silt, pebbles and fine particles of gold and deposited them over a long period of time. Eventually these deposits accumulated and became compacted to form gold-bearing sedimentary rock or **gold reefs**. It is because of this complex, but unique, set of circumstances that South Africa's gold deposits are so concentrated in that area. In other countries like Zimbabwe, gold occurs in smaller 'pockets', which are scattered over a much greater area.

21.5.2 Mining the Gold

A number of different techniques can be used to mine gold. The three most common methods in South Africa are **panning**, **open cast** and **shaft** mining.

1. **Panning**

   Panning for gold is a manual technique that is used to sort gold from other sediments. Wide, shallow pans are filled with sand and gravel (often from river beds) that may contain gold. Water is added and the pans are shaken so that the gold is sorted from the rock and other materials. Because gold is much more dense, it settles to the bottom of the pan. **Pilgrim's Rest** in Mpumalanga, was the first site for gold panning in South Africa.

2. **Open cast mining**

   This is a form of surface mining. Surface layers of rock and sediments are removed so that the deeper gold rich layers can be reached. This type of mining is not suitable if the gold is buried very deep below the surface.

3. **Shaft mining**

   South Africa's thin but extensive gold reefs slope at an angle underneath the ground, and this means that some deposits are very deep and often difficult to reach. Shaft mining is needed to reach the gold ore. After the initial drilling, blasting and equipping of a mine shaft, tunnels are built leading outwards from the main shaft so that the gold reef can be reached. Shaft mining is a dangerous operation, and roof supports are needed so that the rock does not collapse. There are also problems of the intense heat and high pressure below the surface which make shaft mining very complex, dangerous and expensive. A diagram illustrating open cast and shaft mining is shown in figure 21.2.

21.5.3 Processing the gold ore

For every ton of ore that is mined, only a very small amount of gold is extracted. A number of different methods can be used to separate gold from its ore, but one of the more common

Figure 21.2: Diagram showing open cast and shaft mining

methods is called **gold cyanidation**.

In the process of gold cyanidation, the ore is crushed and then cyanide solution is added so that the gold particles are chemically dissolved from the ore. In this step of the process, gold is oxidised. Zinc dust is then added to the cyanide solution. The zinc takes the place of the gold, so that the gold is precipitated out of the solution. This process is shown in figure 21.3.

Interesting Fact

Another method that is used to process gold is called the 'carbon-in-pulp' (CIP) method. This method makes use of the high affinity that activated carbon has for gold, and there are three stages to the process. The first stage involves the *absorption* of gold in the cyanide solution to carbon. In the *elution* stage, gold is removed from the carbon into an alkaline cyanide solution. In the final stage, *electro-winning* is used to remove gold from the solution through a process of electrolysis. Gold that has been removed is deposited on steel wool electrodes. Both the carbon and the acid are then treated so that they can be re-used.

21.5.4 Characteristics and uses of gold

Gold has a number of uses because of its varied and unique characteristics. Below is a list of some of these characteristics that have made gold such a valuable metal:

- *Shiny*

  Gold's beautiful appearance has made it one of the favourite metals for use in jewelery.

STEP 1 - The ore is crushed until it is in fine pieces

↓

STEP 2 - A sodium cyanide (NaCN) solution is mixed with the finely ground rock

$$4Au + 8NaCN + O_2 + 2H_2O \rightarrow 4NaAu(CN)_2 + 4NaOH$$

Gold is **oxidised**.

↓

STEP 3 - The gold-bearing solution can then be separated from the remaining solid ore through filtration

↓

STEP 4 - Zinc is added. Zinc replaces the gold in the gold-cyanide solution. The gold is precipitated from the solution. This is the **reduction** part of the reaction.

Figure 21.3: Flow diagram showing how gold is processed

- *Durable*

  Gold does not tarnish or corrode easily, and therefore does not deteriorate in quality. It is sometimes used in dentistry to make the crowns for teeth.

- *Malleable and ductile*

  Since gold can be bent and twisted into shape, as well as being flattened into very thin sheets, it is very useful in fine wires and to produce sheets of gold.

- *Good conductor*

  Gold is a good conductor of electricity and is therefore used in transistors, computer circuits and telephone exchanges.

- *Heat ray reflector*

  Because gold reflects heat very effectively, it is used in space suits and in vehicles. It is also used in the protective outer coating of artificial satellites. One of the more unusual applications of gold is its use in firefighting, where a thin layer of gold is placed in the masks of the firefighters to protect them from the heat.

Activity :: Case Study : Dropping like a gold balloon

Read the article below, which has been adapted from one that appeared in the *Financial Mail* on 15th April 2005 and then answer the questions that follow.

> As recently as 1980, South Africa accounted for over 70% of world gold production. In 2004, that figure was a dismal 14%. Chamber of Mines figures showed that SA's annual gold production last year slipped to its lowest level since 1931.
>
> Chamber economist Roger Baxter says the 'precipitous' fall in production was caused by the dual impact of the fall in the rand gold price due to the strong rand, and the continued upward rise in costs. Many of these costs, laments Baxter, are 'costs we do not have control over'. These include

water, transport, steel and labour costs, which rose by 13% on average in 2004.

He provides a breakdown of the cost components faced by mines:

- Water prices have risen by 10% per year for the past 3 years
- Steel prices have increased by double-digit rates for each of the past 3 years
- Spoornet's tariffs rose 35% in 2003 and 16.5% in 2004
- Labour costs, which make up 50% of production costs, rose above inflation in 2003 and 2004

At these costs, and at current rand gold prices, about 10 mines, employing 90 000 people, are marginal or loss-making, says Baxter.

1. Refer to the table below showing SA's gold production in tons between 1980 and 2004.

Year	**Production (t)**
1980	675
1985	660
1990	600
1995	525
2000	425
2004	340

   Draw a line graph to illustrate these statistics.
2. What percentage did South Africa's gold production contribute towards global production in:
   (a) 1980
   (b) 2004
3. Outline two reasons for this drop in gold production.
4. Briefly explain how the increased cost of resources such as water contributes towards declining profitability in gold mines.
5. Suggest a reason why the cost of *steel* might affect gold production.
6. Suggest what impact a decrease in gold production is likely to have on...
   (a) South Africa's economy
   (b) mine employees
7. Find out what the current price of gold is. Discuss why you think gold is so expensive.

21.5.5 Environmental impacts of gold mining

However, despite the incredible value of gold and its usefulness in a variety of applications, all mining has an environmental cost. The following are just a few of the environmental impacts of gold mining:

- *Resource consumption*

  Gold mining needs large amounts of electricity and water.

- *Poisoned water*

  Acid from gold processing can leach into nearby water systems such as rivers, causing damage to animals and plants, as well as humans that may rely on that water for drinking. The disposal of other toxic waste (e.g. cyanide) can also have a devastating effect on biodiversity.

- *Solid waste*

  This applies particularly to open pit mines, where large amounts of soil and rock must be displaced in order to access the gold reserves. Processing the gold ore also leaves solid waste behind.

- *Air pollution*

  Dust from open pit mines, as well as harmful gases such as sulfur dioxide and nitrogen dioxide which are released from the furnaces, contribute to air pollution.

- *Threaten natural areas*

  Mining activities often encroach on protected areas and threaten biodiversity in their operation areas.

Activity :: Discussion : Mine rehabilitation

There is a growing emphasis on the need to rehabilitate old mine sites that are no longer in use. If it is too difficult to restore the site to what it was before, then a new type of land use might be decided for that area. Any mine rehabilitation programme should aim to achieve the following:

- ensure that the site is safe and stable
- remove pollutants that are contaminating the site
- restore the biodiversity that was there before mining started
- restore waterways to what they were before mining

There are different ways to achieve these goals. Plants for example, can be used to remove metals from polluted soils and water, and can also help to stabilise the soil so that other vegetation can grow. Land contouring can help to restore drainage in the area.

Discussion:

In groups of 3-4, discuss the following questions:

1. What are the main benefits of mine rehabilitation?
2. What are some of the difficulties that may be experienced in trying to rehabilitate a mine site?
3. Suggest some creative ideas that could be used to encourage mining companies to rehabilitate old sites.
4. One rehabilitation project that has received a lot of publicity is the rehabilitation of dunes that were mined for titanium by Richards Bay Minerals (RBM). As a group, carry out your own research to try to find out more about this project.
   - What actions did RBM take to rehabilitate the dunes?
   - Was the project successful?
   - What were some of the challenges faced?

?

Exercise: Gold mining

Mapungubwe in the Limpopo Province is evidence of gold mining in South Africa as early as 1200. Today, South Africa is a world leader in the technology of gold mining. The following flow diagram illustrates some of the most important steps in the recovery of gold from its ore.

1. Name the process indicated by A.
2. During process A, gold is extracted from the ore. Is gold oxidised or reduced during this process?
3. Use oxidation numbers to explain your answer to the question above.
4. Name the chemical substance that is used in process B.
5. During smelting (illustrated by C in the diagram), gold is sent into a calcining furnace. Briefly explain the importance of this process taking place in the furnace.
6. The recovery of gold can have a negative impact on water in our country, if not managed properly. State at least one negative influence that the recovery of gold can have on water resources and how it will impact on humans and the environment.

21.6 Mining and mineral processing: Iron

Iron is one of the most abundant metals on Earth. Its concentration is highest in the core, and lower in the crust. It is extracted from **iron ore** and is almost never found in its elemental form. Iron ores are usually rich in **iron oxide** minerals and may vary in colour from dark grey to rusty red. Iron is usually found in minerals such as magnetite (Fe_3O_4) and hematite (Fe_2O_3). Iron ore also contains other elements, which have to be removed in various ways. These include silica (Si), phosphorus (P), aluminium (Al) and sulfur (S).

21.6.1 Iron mining and iron ore processing

One of the more common methods of mining for iron ore is **open cast mining**. Open cast mining is used when the iron ore is found near the surface. Once the ore has been removed, it needs to be crushed into fine particles before it can be processed further.

As mentioned earlier, iron is commonly found in the form of **iron oxides**. To create pure iron, the ore must be **smelted** to remove the oxygen.

Definition: Smelting
Smelting is a method used to extract a metal from its ore and then purify it.

Smelting usually involves heating the ore and also adding a reducing agent (e.g. carbon) so that the metal can be freed from its ore. The bonds between iron and oxygen are very strong, and therefore it is important to use an element that will form stronger bonds with oxygen that the iron. This is why carbon is used. In fact, carbon monoxide is the main ingredient that is needed to strip oxygen from iron. These reactions take place in a **blast furnace**.

A blast furnace is a huge steel container many metres high and lined with heat-resistant material. In the furnace the solid raw materials, i.e. iron ore, carbon (in the form of 'coke', a type of coal) and a flux (e.g. limestone) are fed into the top of the furnace and a blast of heated air is forced into the furnace from the bottom. Temperatures in a blast furnace can reach 2000°C. A simple

diagram of a blast furnace is shown in figure 21.4. The equations for the reactions that take place are shown in the flow diagram below.

STEP 1: Production of carbon monoxide

$$C + O_2 \rightarrow CO_2$$

$$CO_2 + C \rightarrow 2CO$$

STEP 2: Reduction of iron oxides takes place in a number of stages to produce iron.

$$3Fe_2O_3 + CO \rightarrow 2Fe_3O_4 + CO_2$$

$$Fe_3O_4 + CO \rightarrow 3FeO + CO_2$$

$$FeO + CO \rightarrow Fe + CO_2$$

STEP 3: Fluxing

The flux is used to melt impurities in the ore. A common flux is limestone ($CaCO_3$). Common impurities are silica, phosphorus (makes steel brittle), aluminium and sulfur (produces SO_2 gases during smelting and interferes with the smellting process).

$$CaCO_3 \rightarrow CaO + CO_2$$

$$CaO + SiO_2 \rightarrow CaSiO_3$$

In step 3, the calcium carbonate breaks down into calcium oxide and carbon dioxide. The calcium oxide then reacts with silicon dioxide (the impurity) to form a **slag**. In this case the slag is the $CaSiO_3$. The slag melts in the furnace, whereas the silicon dioxide would not have, and floats on the more dense iron. This can then be separated and removed.

Figure 21.4: A blast furnace, showing the reactions that take place to produce iron

21.6.2 Types of iron

Iron is the most used of all the metals. Its combination of low cost and high strength make it very important in applications such as industry, automobiles, the hulls of large ships and in the

structural components of buildings. Some of the different forms that iron can take include:

- **Pig iron** is raw iron and is the direct product when iron ore and coke are smelted. It has between 4% and 5% carbon and contains varying amounts of contaminants such as sulfur, silicon and phosphorus. Pig iron is an intermediate step between iron ore, cast iron and steel.
- **Wrought iron** is commercially pure iron and contains less than 0.2% carbon. It is tough, malleable and ductile. Wrought iron does not rust quickly when it is used outdoors. It has mostly been replaced by mild steel for 'wrought iron' gates and blacksmithing. Mild steel does not have the same corrosion resistance as true wrought iron, but is cheaper and more widely available.
- **Steel** is an alloy made mostly of iron, but also containing a small amount of carbon. Elements other than carbon can also be used to make alloy steels. These include manganese and tungsten. By varying the amounts of the alloy elements in the steel, the following characteristics can be altered: hardness, elasticity, ductility and tensile strength.
- **Corrugated iron** is actually sheets of galvanised steel that have been rolled to give them a corrugated pattern. Corrugated iron is a common building material.

One problem with iron and steel is that pure iron and most of its alloys rust. These products therefore need to be protected from water and oxygen, and this is done through painting, galvanisation and plastic coating.

Interesting Fact

Iron is also a very important element in all living organisms. One important role that iron plays is that it is a component of the protein **haemoglobin** which is the protein in blood. It is the iron in the haemoglobin that helps to attract and hold oxygen so that this important gas can be transported around the body in the blood, to where it is needed.

21.6.3 Iron in South Africa

The primary steel industry is an important part of the South African economy and it generates a great deal of foreign exchange.

- About 40 million tons of iron ore is mined annually in South Africa. Approximately 15 million tons are consumed locally, and the remaining 25 million tons are exported.
- South Africa is ranked about 20th in the world in terms of its crude steel production.
- South Africa is the largest steel producer in Africa.
- South Africa exports crude steel and finished steel products, and a lot is also used locally.
- Some of the products that are manufactured in South Africa include reinforcing bars, railway track material, wire rod, plates and steel coils and sheets.

?

Exercise: Iron

Iron is usually extracted from heamatite (iron(III)oxide). Iron ore is mixed with limestone and coke in a blast furnace to produce the metal. The following incomplete word equations describe the extraction process:

A coke + oxygen → gasX

B gasX + coke → gasY

C iron(III)oxide + gasY → iron + gasX

1. Name the gases X and Y.
2. Write a balanced chemical equation for reaction C.
3. What is the function of gas Y in reaction C?
4. Why is limestone added to the reaction mixture?
5. Briefly describe the impact that the mining of iron has on the economy and the environment in our country.

(DoE Exemplar Paper, Grade 11, 2007)

21.7 Mining and mineral processing: Phosphates

A phosphate is a salt of **phosphoric acid** (H_3PO_4). Phosphates are the naturally occurring form of the element phosphorus. Phosphorus is seldom found in its pure elemental form, and **phosphate** therefore refers to a rock or ore that contains phosphate ions. The chemical formula for the phosphate ion is PO_4^{3-}.

21.7.1 Mining phosphates

Phosphate is found in beds in sedimentary rock, and has to be quarried to access the ore. A quarry is a type of open pit mine that is used to extract ore. In South Africa, the main phosphate producer is at the Palaborwa alkaline igneous complex, which produces about 3 million tons of ore per year. The ore is crushed into a powder and is then treated with sulfuric acid to form a superphosphate ($Ca(H_2PO_4)_2$), which is then used as a fertilizer. In the equation below, the phosphate mineral is calcium phosphate ($Ca_3(PO_4)_2$.

$$Ca_3(PO_4)_2 + 2H_2SO_4 \rightarrow Ca(H_2PO_4)_2 + 2CaSO_4$$

Alternatively, the ore can be treated with concentrated phosphoric acid, in which case the reaction looks like this:

$$3Ca_3(PO_4)_2.CaF_2 + 4H_3PO_4 + 9H_2O \rightarrow 9Ca(H_2PO_4)_2 + CaF_2$$

21.7.2 Uses of phosphates

Phosphates are mostly used in **agriculture**. Phosphates are one of the three main nutrients needed by plants, and they are therefore an important component of **fertilisers** to stimulate plant growth.

Interesting Fact

Exploring the lithosphere for minerals is not a random process! **Geologists** help to piece together a picture of what past environments might have been like, so that predictions can be made about where minerals might have a high concentration. **Geophysicists** measure gravity, magnetics and the electrical properties of rocks to try to make similar predictions, while **geochemists** sample the soils at the earth's surface to get an idea of what lies beneath them. You can see now what an important role scientists play in mineral exploration!

Exercise: Phosphates

Rock phosphate [$Ca_{10}(PO_4)_6F_2$], mined from open pit mines at Phalaborwa, is an important raw material in the production of fertilisers. The following two reactions are used to transform rock phosphate into water soluble phosphates:

A: $Ca_{10}(PO_4)_6F_2 + 7X + 3H_2O \rightarrow 3Ca(H_2PO_4)_2H_2O + 2HF + 7CaSO_4$

B: $Ca_{10}(PO_4)_6F_2 + 14Y + 10H_2O \rightarrow 10Ca(H_2PO_4)_2H_2O + 2HF$

1. Identify the acids represented by X and Y.
2. Despite similar molecular formulae, the products $Ca(H_2PO_4)_2$ formed in the two reactions have different common names. Write down the names for each of these products in reactions A and B.
3. Refer to the products in reactions A and B and write down TWO advantages of reaction B over reaction A.
4. Why is rock phosphate unsuitable as fertiliser?
5. State ONE advantage and ONE disadvantage of phosphate mining.

(DoE Exemplar Paper, Grade 11, 2007)

Activity :: Case Study : Controversy on the Wild Coast - Titanium mining

Read the extract below, which has been adapted from an article that appeared in the Mail and Guardian *on 4th May 2007, and then answer the questions that follow.*

> A potentially violent backlash looms in Pondoland over efforts by an Australian company to persuade villagers to back controversial plans to mine an environmentally sensitive strip of the Wild Coast. The mining will take place in the Xolobeni dunes, south of Port Edward. The application has outraged environmental groups, largely because the proposed mining areas form part of the Pondoland centre of endemism, which has more species than the United Kingdom, some of which are endemic and facing extinction.
>
> Exploratory drilling revealed Xolobeni has the world's 10th largest titanium deposit, worth about R11 billion. The amount of money that will be spent over the mine's 22 years, including a smelter, is estimated at R1.4 billion. The Australian mining company predicts that 570 direct jobs will be created.
>
> But at least two communities fiercely oppose the mining plans. Some opponents are former miners who fear Gauteng's mine dumps and compounds will be replicated on the Wild Coast. Others are employees of failed ecotourism ventures, who blame the mining company for their situation. Many are suspicious of outsiders. The villagers have also complained that some of the structures within the mining company are controlled by business leaders with political connections, who are in it for their own gain. Intimidation of people who oppose the mining has also been alleged. Headman Mandoda Ndovela was shot dead after his outspoken criticism of the mining.
>
> Mzamo Dlamini, a youth living in one of the villages that will be affected by the mining, said 10% of the Amadiba 'who were promised riches by the mining company' support mining. 'The rest say if people bring those machines, we will fight.'

1. Explain what the following words means:
   (a) endemic
   (b) smelter
   (c) ecotourism
2. What kinds of 'riches' do you think the Amadiba people have been promised by the mining company?
3. In two columns, list the potential **advantages** and **disadvantages** of mining in this area.
4. Imagine that you were one of the villagers in this area. Write down *three questions* that you would want the mining company to answer before you made a decision about whether to oppose the mining or not. Share your ideas with the rest of the class.
5. Imagine that you are an environmentalist. What would your main concerns be about the proposed mining project? Again share your answers with the rest of the class.

21.8 Energy resources and their uses: Coal

The products of the lithosphere are also important in meeting our **energy needs**. Coal is one product that is used to produce energy. In South Africa, coal is particularly important because most of our electricity is generated using coal as a fuel. South Africa is the world's sixth largest coal producer, with Mpumalanga contributing about 83% of our total production. Other areas in which coal is produced, include the Free State, Limpopo and KwaZulu-Natal. One of the problems with coal however, is that it is a **non-renewable resource**, meaning that once all resources have been used up, it cannot simply be produced again. Burning coal also produces large quantities of greenhouse gases, which may play a role in global warming. At present, ESKOM, the South African government's electric power producer, is the coal industry's main customer.

21.8.1 The formation of coal

Coal is what is known as a **fossil fuel**. A fossil fuel is a *hydrocarbon* that has been formed from organic material such as the remains of plants and animals. When plants and animals decompose, they leave behind organic remains that accumulate and become compacted over millions of years under sedimentary rock. Over time, the *heat* and *pressure* in these parts of the earth's crust also increases, and coal is formed. When coal is burned, a large amount of heat energy is released, which is used to produce electricity. **Oil** is also a fossil fuel and is formed in a similar way.

Definition: Fossil Fuel
A fossil fuel is a hydrocarbon that is formed from the fossilised remains of dead plants and animals that have been under conditions of intense heat and pressure for millions of years.

21.8.2 How coal is removed from the ground

Coal can be removed from the crust in a number of different ways. The most common methods used are *strip mining*, *open cast mining* and *underground mining*.

1. **Strip mining**

   Strip mining is a form of surface mining that is used when the coal reserves are very shallow. The *overburden* (overlying sediment) is removed so that the coal seams can be reached. These sediments are replaced once the mining is finished, and in many cases, attempts are made to *rehabilitate* the area.

2. **Open cast mining**

   Open cast mining is also a form of surface mining, but here the coal deposits are too deep to be reached using strip mining. One of the environmental impacts of open cast mining is that the overburden is dumped somewhere else away from the mine, and this leaves a huge pit in the ground.

3. **Underground mining**

   Undergound mining is normally used when the coal seams are amuch deeper, usually at a depth greater than 40 m. As with shaft mining for gold, the problem with underground mining is that it is very dangerous, and there is a very real chance that the ground could collapse during the mining if it is not supported. One way to limit the danger is to use *pillar support* methods, where some of the ground is left unmined so that it forms pillars to support the roof. All the other surfaces underground will be mined. Using another method called *longwalling*, the roof is allowed to collapse as the mined-out area moves along. In South Africa, only a small percentage of coal is mined in this way.

21.8.3 The uses of coal

Although in South Africa, the main use of coal is to produce electricity, it can also be used for other purposes.

1. **Electricity**

   In order to generate electricity, solid coal must be crushed and then burned in a furnace with a boiler. A lot of steam is produced and this is used to spin turbines which then generate electricity.

2. **Gasification**

   If coal is broken down and subjected to very high temperatures and pressures, it forms a *synthesis gas*, which is a mix of carbon dioxide and hydrogen gases. This is very important in the *chemical industry* (refer to chapter 23).

3. **Liquid fuels**

   Coal can also be changed into liquid fuels like petrol and diesel using the **Fischer-Tropsch process**. In fact, South Africa is one of the leaders in this technology (refer to chapter 23). The only problem is that producing liquid fuels from coal, rather than refining petroleum that has been drilled, releases much greater amounts of carbon dioxide into the atmosphere, and this contributes further towards global warming.

21.8.4 Coal and the South African economy

In South Africa, the coal industry is second only to the gold industry. More than this, South Africa is one of the world's top coal *exporters*, and also one of the top *producers*. Of the coal that is produced, most is used locally to produce electricity and the rest is used in industry and domestically.

The problem with coal however, is that it is a **non-renewable resource** which means that once all the coal deposits have been mined, there will be none left. Coal takes such a long time to form, and requires such specific environmental conditions, that it would be impossible for coal to re-form at a rate that would keep up with humankind's current consumption. It is therefore very important that South Africa, and other countries that rely on coal, start to look for alternative energy resources.

21.8.5 The environmental impacts of coal mining

There are a number of environmental impacts associated with coal mining.

- *Visual impact and landscape scars*

  Coal mining leaves some very visible scars on the landscape, and destroys biodiversity (e.g. plants, animals). During strip mining and open cast mining, the visual impact is particularly bad, although this is partly reduced by rehabilitation in some cases.

- *Spontaneous combustion and atmospheric pollution*

  Coal that is left in mine dumps may spontaneously combust, producing large amounts of sulfurous smoke which contributes towards atmospheric pollution.

- *Acid formation*

  Waste products from coal mining have a high concentration of sulfur compounds. When these compounds are exposed to water and oxygen, sulfuric acid is formed. If this acid washes into nearby water systems, it can cause a lot of damage to the ecosystem. Acid can also leach into soils and alter its acidity. This in turn affects what will be able to grow there.

- *Global warming*

  As was discussed earlier, burning coal to generate electricity produces *carbon dioxide* and *nitrogen oxides* which contribute towards global warming (refer to chapter 22). Another gas that causes problems is *methane*. All coal contains methane, and deeper coal contains the most methane. As a greenhouse gas, methane is about twenty times more potent than carbon dioxide.

Interesting Fact

It is easy to see how mining, and many other activities including industry and vehicle transport, contribute towards Global Warming. It was for this reason that South Africa joined the **Carbon Sequestration Leadership Forum (CSLF)**. The forum is an international climate change initiative that focuses on developing cost effective technologies to separate and capture carbon dioxide from the atmosphere so that it can be stored in some way. The CSLF also aims to make these technologies as widely available as possible.

Exercise: Coal in South Africa

The following advertisement appeared in a local paper:

1. *"Coal is as old as the hills, and just as natural."* Is this statement TRUE? Motivate your answer by referring to how coal is formed.
2. Coal is a non-renewable energy source. Quote a statement from the advertisement that gives an indication that coal is non-renewable. Give a reason for your choice.
3. Is coal actually a healthy source of energy? Motivate your answer by referring to all influences that coal and coal mining have on both humans and the environment.
4. Why is coal used as a primary energy source in South Africa?

(DoE Exemplar Paper 2, Grade 11, 2007)

21.9 Energy resources and their uses: Oil

Oil is another product of the lithosphere which is very important in meeting our fuel needs.

21.9.1 How oil is formed

Oil is formed in a very similar way to coal, except that the organic material is laid down in **oceans**. Organisms such as zooplankton and algae settle to the ocean floor and become buried under layers of mud. Over time, as these layers of sediment accumulate and the heat and pressure also increase, the organic material changes to a waxy material called *kerogen*. Eventually, with continuing heat and pressure, liquid and gas **hydrocarbons** are formed. These hydrocarbons are lighter than rock and therefore move upwards through the rock layers before being trapped by an impermeable layer. Here the oil will slowly accumulate until there is enough that it can be accessed by oil rigs and other equipment. **Crude oil** or **petroleum**, is actually a mixture of hydrocarbons (mostly alkanes) of different lengths, ranging from 5 carbons to 18 carbons in the hydrocarbon chain. If the mixture contains mostly short hydrocarbons, then it is a gas called **natural gas**. As the hydrocarbon chains in the mixture become longer, the product becomes more and more solid. Coal is made up of the longest hydrocarbons. For more information on hydrocarbons, refer to chapter 9.

21.9.2 Extracting oil

When enough oil has accumulated in a well, it becomes economically viable to try to extract it either through **drilling** or **pumping**. If the pressure in the oil reservoir is high, the oil is forced

naturally to the surface. This is known as *primary recovery* of oil. If the pressure is low, then pumps must be used to extract it. This is known as *secondary recovery*. When the oil is very difficult to extract, steam injection into the reservoir can be used to heat the oil, reduce its viscosity and make it easier to extract.

While most of South Africa's oil is imported and then processed at a refinery in either Durban, Cape Town or Sasolburg, some is extracted from coal, as discussed in section 21.8.

21.9.3 Other oil products

Oil can also be used to make a variety of different products. You will find more information on this in chapter 23.

- *Fractional distillation*

  Fractional distillation is the separation of a mixture into the parts that make it up. In oil refineries, crude oil is separated into useful products such as asphalt, diesel, fuel oil, gasoline, kerosine, liquid petroleum gas (LPG) and tar, to name just a few.

- *Cracking*

  There are two types of cracking, *steam cracking* and *hydrocracking*. Cracking is used to change heavy hydrocarbons such as petroleum into lighter hydrocarbons such as fuels (LPG and gasoline), plastics (ethylene) and other products that are needed to make fuel gas (propylene).

21.9.4 The environmental impacts of oil extraction and use

Some of the key environmental impacts associated with the extraction and use of oil are as follows:

- *Pollution*

  Exploring the oceans for oil, and the actual drilling process, can result in major pollution.

- *Ecosystem impacts*

  Dredging the ocean floors for oil can disrupt seabed ecosystems.

- *Global warming*

  Burning oil as a fuel source produces carbon dioxide, which contributes towards global warming.

21.10 Alternative energy resources

As the world's population increases, so does the demand for energy. As we have already mentioned, many of our energy resources are **non-renewable** and will soon run out. In addition, many of the fuels that we use produce large amounts of greenhouse gases, which can contribute towards global warming. If we are to maintain the quality and health of our planet, and also meet our growing need for energy, we will need to investigate alternative energy resources. In this next section, we are going to take a closer look at some of these possible alternatives. Many of these options are very controversial, and may have both pros and cons.

- *Nuclear power*

  Another element that is found in the crust, and which helps to meet our energy needs, is **uranium**. Uranium produces energy through the process of *nuclear fission* (chapter **??**). Neutrons are aimed at the nucleii of the uranium atoms in order to split them. When the nucleus of a uranium atom is split, a large amount of energy is released as heat. This heat is used to produce steam, which turns turbines to generate electricity. Uranium is

produced as a by-product of gold in some mines in the Witwatersrand, and as a by-product in some copper mines, for example in Palaborwa. Many people regard this type of nuclear power as relatively environmentally friendly because it doesn't produce a lot of greenhouse gases. However, generating nuclear power does produce radioactive wastes, which must be carefully disposed of in order to prevent contamination. There are also concerns around leaking of nuclear materials.

- *Natural gas*

  Natural gas is formed in a similar way to oil and is often located above oil deposits in the earth's crust. 'Natural gas' refers to a hydrocarbon gas, composed mostly of methane. It is highly combustible and produces low emissions.

  In June 2002, construction began on a pipeline that would stretch for 865 km between Mozambique and South Africa. Mozambique has large sources of under-utilised natural gas and so an agreement was reached between SASOL and the South African and Mozambican governments to build the pipeline, which would transport natural gas from Mozambique to South Africa. The benefits of natural gas include the fact that it is a clean-burning fossil fuel and few by-products are emitted as pollutants. It is also an economical and efficient energy source as the gas can easily be piped directly to a customer's facility.

- *Biofuels*

  In many parts of the world, ethanol is currently being used as a substitute for crude petroleum. Ethanol can be produced through the fermentation of sugar-containing products such as sugar cane. One of the problems with this however, is the vast areas of land that are needed to cultivate the necessary crops. Crops such as maize can also be used in the process. In South Africa, a company called 'Ethanol Africa' has been set up by commercial farmers to convert their surplus maize into environmentally-friendly biofuel, and plans are underway to establish ethanol plants in some of the maize-producing provinces.

- *Hydropower*

  Hydropower produces energy from the action of falling water. As water falls from a height, the energy is used to turn turbines which produce electricity. However, for hydropower to be effective, a large dam is needed to store water. The building of a dam comes with its own set of problems such as the expense of construction, as well as the social and environmental impacts of relocating people (if the area is populated),and disrupting a natural river course.

- *Solar energy*

  Solar energy is energy from the sun. The sun's radiation is trapped in solar panels and is then converted into electricity. While this process is environmentally friendly, and solar energy is a renewable resource, the supply of radiation is not constant (think for example of cloudy days, and nights), and the production of electricity is not efficient. Solar energy can however meet small energy needs such as the direct heating of homes.

- *Geothermal energy*

  This type of energy comes from the natural heat below the Earth's surface. If hot underground steam can be tapped and brought to the surface, it has the potential to produce electricity.

Activity :: Discussion : Using energy wisely

The massive power cuts or 'load shedding' that South Africans began to experience at the beginning of 2008, were a dramatic wake-up call to the growing energy crisis that the country faces.

There are alternative energy sources available, but they will take years to become functional, and many of them have their own problems. Another way to look at the problem, is to put the focus on reducing how much energy is *used* rather than focusing only on ways to meet the growing demand.

1. In your groups, discuss ways that each of the following groups of people could save energy.
   (a) industries
   (b) domestic users
   (c) farmers
2. Discuss creative incentives that could be used to encourage each of these groups to reduce their energy consumption.

21.11 Summary

- The **lithosphere** is the solid, outermost part of our planet and contains many important metal elements such as gold and iron, as well as products that are needed to produce energy.
- These elements are seldom found in their pure form, but rather as **minerals** in **rocks**.
- A **mineral** is formed through geological processes. It can either be a pure element (e.g. gold) or may consist of a number of different elements e.g. the gold-bearing mineral calaverite ($AuTe_2$).
- A **rock** is an aggregate of a number of different minerals.
- An **ore** is rock that contains minerals which make it valuable for mining.
- Minerals have been used throughout **history**. As new metals and minerals were discovered, important growth took place in industry, agriculture and technology.
- **Gold** is one of the most important metals in the history of South Africa. It was the discovery of gold that led to an influx of fortune-seeking foreigners, and a growth in mining villages and towns.
- Most of South Africa's gold is concentrated in the 'Golden Arc' in the area between Johannesburg and Welkom.
- Three methods are used to obtain gold from the lithosphere: **panning**, **open cast mining** and **shaft mining**.
- Gold ore must be processed so that the metal can be removed. One way to process the ore after it has been crushed is a method called **gold cyanidation**. A cyanide solution is added to the crushed ore so that a gold-cyanide solution is formed. Zinc is then added to this solution so that the gold is precipitated out.
- Gold has a number of important **characteristics** which make it a useful metal for jewelery and other applications. The metal is shiny, durable, malleable, ductile, is a good conductor of electricity and is also a good heat reflector.
- Gold mining has a number of **environmental impacts**, which include resource consumption, air pollution, poisoned water, solid waste and the destruction of biodiversity in natural areas.
- **Mine rehabilitation** is one way of restoring old mine sites to what they were like before.
- **Iron** is another important metal and is used in industry, furniture and building materials.
- Iron is usually found in minerals such as **iron oxides**. These minerals must be processed to remove the metal.
- When iron ore is **processed**, a blast furnace is used. The iron ore, carbon and a flux are added to the top of the furnace and hot air is blasted into the bottom of the furnace. A number of reactions occur in the furnace to finally remove the iron from its ore. Iron oxides are reduced by carbon monoxide to produce iron.

- Iron can occur in a number of forms, depending on its level of purity and carbon content. It can also occur in an **alloy** e.g. steel.
- **Phosphates** are found in sedimentary rock, which must be quarried to access the ore.
- Phosphates react with phosphoric acid or sulfuric acid to produce a **superphosphate** ($Ca(H_2PO_4)_2$), which is an important component in **fertilisers**.
- The products of the lithosphere are also important in meeting **energy needs**. **Coal** and **oil** can be extracted from the lithosphere for this purpose.
- Coal and oil are both **fossil fuels**. A fossil fuel is a hydrocarbon that has been formed from the fossilsed remains of plants and animals that have been under conditions of high heat and pressure over a long period of time.
- Coal and oil are **non-renewable resources**, meaning that once they have been used up, no more can be produced.
- Coal can be removed from the ground using **strip mining**, **open cast mining** or **underground mining**.
- Coal is burned to produce energy, which can be used to generate **electricity**. Coal can also be used to produce **liquid fuels** or a **syngas** which can be converted into other useful products for the chemical industry.
- Some of the **environmental impacts** associated with coal mining include landscape scars, spontaneous combustion, acid formation and global warming.
- Oil is also a fossil fuel but it forms in the **oceans**. It can extracted using either **pumping** or **drilling**, depending on the pressure of the oil.
- **Fractional distillation** of oil can be used to make products such as diesel, gasoline and liquid petroleum gas.
- **Cracking** can be used to convert heavy hydrocarbons to light hydrocarbons.
- The **environmental impacts** of oil extraction and use are similar to those for coal.
- In view of the number of environmental impacts associated with the extraction and use of coal and oil, other **alternative energy sources** should be considered. These include nuclear power, biofuels, hydropower and a number of others. All of these alternatives have their own advantages and disadvantages.

?

Exercise: Summary Exercise

1. Give one word to describe each of the following phrases:
   (a) earth's crust together with the upper layer of the mantle
   (b) a mineral containing silica and oxygen
   (c) an alloy of iron and tin
   (d) a manual technique used to sort gold from other sediments
2. For each of the following questions, choose the *one correct answer* from the list provided.
   (a) One of the main reasons that South Africa's gold industry has been so economically viable is that...
      i. gold panning can be used as an additional method to extract gold
      ii. open cast mining can be used to extract gold reserves
      iii. South Africa's geological history is such that its gold reserves are concentrated in large reefs
      iv. South Africa has large amounts of water to use in mining
   (b) The complete list of reactants in an iron blast furnace is...

i. carbon and oxygen
ii. coal, oxygen, iron ore and limestone
iii. carbon, oxygen and iron ore
iv. coal, air, iron ore and slag

3. **More profits, more poisons**
In the last three decades, gold miners have made use of *cyanidation* to recover gold from the ore. Over 99% of gold from ore can be extracted in this way. It allows miners to obtain gold flakes - too small for the eye to see. Gold can also be extracted from the waste of old operations which sometimes leave as much as a third of the gold behind.
The left-over cyanide can be re-used, but is more often stored in a pond behind a dam or even dumped directly into a local river. A teaspoonful of 2% solution of cyanide can kill a human adult.
Mining companies insist that cyanide breaks down when exposed to sunlight and oxygen which render it harmless. They also point to scientific studies that show that cyanide swallowed by fish will not 'bio-accumulate', which means it does not pose a risk to anyone who eats the fish. In practice, cyanide solution that seeps into the ground will not break down because of the absence of sunlight. If the cyanide solution is very acidic, it could turn into cyanide gas, which is toxic to fish. On the other hand, if the solution is alkaline the cyanide does not break down.
There are no reported cases of human death from cyanide spills. If you don't see corpses, everything is okay.

(a) What is *cyanidation*?
(b) What type of chemical reaction takes place during this process: precipitation, acid-base or redox?
(c) Is the pH of the solution after cyanidation greater than, less than or equal to 7?
(d) How is solid gold recovered from this solution?
(e) Refer to cyanidation and discuss the meaning of the heading of this extract: *More profits, more poisons*.
(DoE Grade 11 Paper 2, 2007)

Chapter 22

The Atmosphere - Grade 11

Our earth is truly an amazing planet! Not only is it exactly the right distance from the sun to have temperatures that will support life, but it is also one of the only planets to have liquid water on its surface. In addition, our earth has an atmosphere that has just the right composition to allow life to exist. The **atmosphere** is the layer of gases that surrounds the earth. We may not always be aware of them, but without these gases, life on earth would definitely not be possible. The atmosphere provides the gases that animals and plants need for respiration (breathing) and photosynthesis (the production of food), it helps to keep temperatures on earth constant and also protects us from the sun's harmful radiation.

In this chapter, we are going to take a closer look at the chemistry of the earth's atmosphere and at some of the human activities that threaten the delicate balance that exists in this part of our planet.

22.1 The composition of the atmosphere

Earth's atmosphere is a mixture of gases. Two important gases are nitrogen and oxygen, which make up about 78.1% and 20.9% of the atmosphere respectively. A third gas, Argon, contributes about 0.9%, and a number of other gases such as carbon dioxide, methane, water vapour, helium and ozone make up the remaining 0.1%. In an earlier chapter, we discussed the importance of nitrogen as a component of proteins, the building blocks of life. Similarly, oxygen is essential for life because it is the gas we need for respiration. We will discuss the importance of some of the other gases later in this chapter.

Interesting Fact

The earth's early atmosphere was very different from what it is today. When the earth formed around 4.5 billion years ago, there was probably no atmosphere. Some scientists believe that the earliest atmosphere contained gases such as water vapour, carbon dioxide, nitrogen and sulfur which were released from inside the planet as a result of volcanic activity. Many scientists also believe that the first stage in the evolution of life, around 4 billion years ago, needed an oxygen-free environment. At a later stage, these primitive forms of plant life began to release small amounts of oxygen into the atmosphere as a product of photosynthesis. During photosynthesis, plants use carbon dioxide, water and sunlight to produce simple sugars. Oxygen is also released in the process.

$$6CO_2 + 6H_2O + \text{sunlight} \rightarrow C_6H_{12}O_6 + 6O_2$$

This build-up of oxygen in the atmosphere eventually led to the formation of the ozone layer, which helped to filter the sun's harmful UV radiation so that plants were able to flourish

in different environments. As plants became more widespread and photosythesis increased, so did the production of oxygen. The increase in the amount of oxygen in the atmosphere would have allowed more forms of life to exist.

If you have ever had to climb to a very high altitude (altitude means the 'height' in the atmosphere), you will have noticed that it becomes very difficult to breathe, and many climbers suffer from 'altitude sickness' before they reach their destination. This is because the density of gases becomes less as you move higher in the atmosphere. It is **gravity** that holds the atmosphere close to the earth. As you move higher, this force weakens slightly and so the gas particles become more spread out. In effect, when you are at a high altitude, the gases in the atmosphere haven't changed, but there are fewer oxygen molecules in the same amount of air that you are able to breathe.

Definition: Earth's atmosphere
The Earth's atmosphere is a layer of gases that surround the planet, and which are held there by the Earth's gravity. The atmosphere contains roughly 78.1% nitrogen, 20.9% oxygen, 0.9% argon, 0.038% carbon dioxide, trace amounts of other gases, and a variable amount of water vapour. This mixture of gases is commonly known as air. The atmosphere protects life on Earth by absorbing ultraviolet solar radiation and reducing temperature extremes between day and night.

22.2 The structure of the atmosphere

The earth's atmosphere is divided into different layers, each with its own particular characteristics (figure 22.1).

22.2.1 The troposphere

The **troposphere** is the lowest level in the atmosphere, and it is the part in which we live. The troposphere varies in thickness, and extends from the ground to a height of about 7km at the poles and about 18km at the equator. An important characteristic of the troposphere is that its temperature *decreases* with an increase in altitude. In other words, as you climb higher, it will get colder. You will have noticed this if you have climbed a mountain, or if you have moved from a city at a high altitude to one which is lower; the average temperature is often lower where the altitude is higher. This is because the troposphere is heated from the 'bottom up'. In other words, places that are closer to the Earth's surface will be warmer than those at higher altitudes. The heating of the atmosphere will be discussed in more detail later in this chapter.

The word troposphere comes from the Greek *tropos*, meaning *turning* or *mixing*. The troposphere is the most turbulent part of the atmosphere and is the part where our **weather** takes place. Weather is the state of the air at a particular place and time e.g. if it is warm or cold, wet or dry, and how cloudy or windy it is. Generally, jet aircraft fly just above the troposphere to avoid all this turbulence.

22.2.2 The stratosphere

Above the troposphere is another layer called the **stratosphere**, where most long distance aircraft fly. The stratosphere extends from altitudes of 10 to 50km. If you have ever been in an aeroplane and have looked out the window once you are well into the flight, you will have noticed

Figure 22.1: A generalised diagram showing the structure of the atmosphere to a height of 110 km

that you are actually flying above the level of the clouds. As we have already mentioned, clouds and weather occur in the troposphere, whereas the stratosphere has very stable atmospheric conditions and very little turbulence. It is easy to understand why aircraft choose to fly here!

The stratosphere is different from the troposphere because its temperature *increases* as altitude increases. This is because the stratosphere absorbs solar radiation directly, meaning that the upper layers closer to the sun will be warmer. The upper layers of the stratosphere are also warmer because of the presence of the **ozone layer**. Ozone (O_3) is formed when solar radiation splits an oxygen molecule (O_2) into two atoms of oxygen. Each individual atom is then able to combine with an oxygen molecule to form ozone. The two reactions are shown below:

$$O_2 \rightarrow O + O$$

$$O + O_2 \rightarrow O_3$$

The change from one type of molecule to another produces energy, and this contributes to higher temperatures in the upper part of the stratosphere. An important function of the ozone layer is to absorb UV radiation and reduce the amount of harmful radiation that reaches the Earth's surface.

Extension: CFCs and the ozone layer

You may have heard people talking about 'the hole in the ozone layer'. What do they mean by this and do we need to be worried about it?

Most of the earth's ozone is found in the stratosphere and this limits the amount of UV radiation that reaches the earth. However, human activities have once again disrupted the chemistry of the atmosphere. Chlorofluorocarbons (CFC's) are compounds found in aerosol cans, fridges and airconditioners. In aerosol cans, it is the CFC's that cause the substance to be sprayed outwards. The bad side of CFC's is that, when they are released into the atmosphere, they break down ozone molecules so that the ozone is no longer able to protect us as much from UV rays. The 'ozone hole' is actually a thinning of the ozone layer approximately above Antarctica. Let's take a closer look at the chemical reactions that are involved in breaking down ozone:

1. When CFC's react with UV radiation, a carbon-chlorine bond in the chlorofluorocarbon breaks and a new compound is formed, with a chlorine atom.

$$CFCl_3 + UV \rightarrow CFCl_2 + Cl$$

2. The single chlorine atom reacts with ozone to form a molecule of chlorine monoxide and oxygen gas. In the process, ozone is destroyed.

$$Cl + O_3 \rightarrow ClO + O_2$$

3. The chlorine monoxide then reacts with a free oxygen atom (UV radiation breaks O_2 down into single oxygen atoms) to form oxygen gas and a single chlorine atom.

$$ClO + O \rightarrow Cl + O_2$$

4. The chlorine atom is then free to attack more ozone molecules, and the process continues. A single CFC molecule can destroy 100 000 ozone molecules.

One possible consequence of ozone depletion is an increase in the incidence of skin cancer because there is more UV radiation reaching earth's surface. CFC replacements are now being used to reduce emissions, and scientists are trying to find ways to restore ozone levels in the atmosphere.

22.2.3 The mesosphere

The mesosphere is located about 50-80/85km above Earth's surface. Within this layer, temperature decreases with increasing altitude. Temperatures in the upper mesosphere can fall as low as -100°C in some areas. Millions of meteors burn up daily in the mesosphere because of collisions with the gas particles that are present in this layer. This leads to a high concentration of iron and other metal atoms.

22.2.4 The thermosphere

The thermosphere exists at altitudes above 80 km. In this part of the atmosphere, ultraviolet (UV) and shorter X-Ray radiation from the sun cause neutral gas atoms to be *ionised*. At these radiation frequencies, photons from the solar radiation are able to dislodge electrons from neutral atoms and molecules during a collision. A *plasma* is formed, which consists of negative free electrons and positive ions. This part of the atmosphere is called the **ionosphere**. At the same time that ionisation takes place however, an opposing process called recombination also begins. Some of the free electrons are drawn to the positive ions, and combine again with them if they are in close enough contact. Since the gas density increases at lower altitudes, the recombination process occurs more often here because the gas molecules and ions are closer together. The

ionisation process produces energy which means that the upper parts of the thermosphere, which are dominated by ionisation, have a higher temperature than the lower layers where recombination takes place. Overall, temperature in the thermosphere increases with an increase in altitude.

Extension: The ionosphere and radio waves

The ionosphere is of practical importance because it allows **radio waves** to be transmitted. A radio wave is a type of electromagnetic radiation that humans use to transmit information without wires. When using high-frequency bands, the ionosphere is used to reflect the transmitted radio beam. When a radio wave reaches the ionosphere, the electric field in the wave forces the electrons in the ionosphere into oscillation at the same frequency as the radio wave. Some of the radio wave energy is given up to this mechanical oscillation. The oscillating electron will then either recombine with a positive ion, or will re-radiate the original wave energy back downward again. The beam returns to the Earth's surface, and may then be reflected back into the ionosphere for a second bounce.

Interesting Fact

The ionosphere is also home to the **auroras**. Auroras are caused by the collision of charged particles (e.g. electrons) with atoms in the earth's upper atmosphere. Charged particles are energised and so, when they collide with atoms, the atoms also become energised. Shortly afterwards, the atoms emit the energy they have gained, as light. Often these emissions are from oxygen atoms, resulting in a greenish glow (wavelength 557.7 nm) and, at lower energy levels or higher altitudes, a dark red glow (wavelength 630 nm). Many other colours can also be observed. For example, emissions from atomic nitrogen are blue, and emissions from molecular nitrogen are purple. Auroras emit visible light (as described above), and also infra-red, ultraviolet and x-rays, which can be observed from space.

Exercise: The composition of the atmosphere

1. Complete the following summary table by providing the missing information for each layer in the atmosphere.

Atmospheric layer	Height (km)	Gas composition	General characteristics
Troposphere	0-18		Turbulent; part of atmosphere where weather occurs
			Ozone reduces harmful radiation reaching Earth
Mesosphere			High concentration of metal atoms
	more than 80 km		

2. Use your knowledge of the atmosphere to explain the following statements:

(a) Athletes who live in coastal areas need to acclimatise if they are competing at high altitudes.
(b) Higher incidences of skin cancer have been recorded in areas where the ozone layer in the atmosphere is thin.
(c) During a flight, turbulence generally decreases above a certain altitude.

22.3 Greenhouse gases and global warming

22.3.1 The heating of the atmosphere

As we mentioned earlier, the distance of the earth from the sun is not the only reason that temperatures on earth are within a range that is suitable to support life. The composition of the atmosphere is also critically important.

The earth receives electromagnetic energy from the sun in the *visible spectrum*. There are also small amounts of infrared and ultraviolet radiation in this incoming solar energy. Most of the radiation is *shortwave* radiation, and it passes easily through the atmosphere towards the earth's surface, with some being reflected before reaching the surface. At the surface, some of the energy is absorbed, and this heats up the earth's surface. But the situation is a little more complex than this.

A large amount of the sun's energy is re-radiated from the surface back into the atmosphere as **infrared** radiation, which is invisible. As this radiation passes through the atmosphere, some of it is absorbed by **greenhouse gases** such as carbon dioxide, water vapour and methane. These gases are very important because they re-emit the energy back towards the surface. By doing this, they help to warm the lower layers of the atmosphere even further. It is this 're-emission' of heat by greenhouse gases, combined with surface heating and other processes (e.g. conduction and advection) that maintain temperatures at exactly the right level to support life. Without the presence of greenhouse gases, most of the sun's energy would be lost and the Earth would be a lot colder than it is! A simplified diagram of the heating of the atmosphere is shown in figure 22.2.

22.3.2 The greenhouse gases and global warming

Many of the greenhouse gases occur naturally in small quantities in the atmosphere. However, human activities have greatly increased their concentration, and this has led to a lot of concern about the impact that this could have in *increasing* global temperatures. This phenomenon is known as **global warming**. Because the natural concentrations of these gases are low, even a small increase in their concentration as a result of human emissions, could have a big effect on temperature. But before we go on, let's look at where some of these human gas emissions come from.

- **Carbon dioxide** (CO_2)

  Carbon dioxide enters the atmosphere through the burning of fossil fuels (oil, natural gas, and coal), solid waste, trees and wood products, and also as a result of other chemical reactions (e.g. the manufacture of cement). Carbon dioxide can also be *removed* from the atmosphere when it is absorbed by plants during photosynthesis.

- **Methane** (CH_4)

  Methane is emitted when coal, natural gas and oil are produced and transported. Methane emissions can also come from livestock and other agricultural practices and from the decay of organic waste.

Figure 22.2: The heating of the Earth's atmosphere

- **Nitrous oxide** (N_2O)

  Nitrous oxide is emitted by agriculture and industry, and when fossil fuels and solid waste are burned.

- **Fluorinated gases** (e.g. hydrofluorocarbons, perfluorocarbons, and sulfur hexafluoride)

  These gases are all *synthetic*, in other words they are man-made. They are emitted from a variety of industrial processes. Fluorinated gases are sometimes used in the place of other ozone-depleting substances (e.g. CFC's). These are very powerful greenhouse gases, and are sometimes referred to as High Global Warming Potential gases ('High GWP gases').

Overpopulation is a major problem in reducing greenhouse gas emissions, and in slowing down global warming. As populations grow, their demands on resources (e.g. energy) increase, and so does their production of greenhouse gases.

Extension: Ice core drilling - Taking a look at earth's past climate

Global warming is a very controversial issue. While many people are convinced that the increase in average global temperatures is directly related to the increase in atmospheric concentrations of carbon dioxide, others argue that the climatic changes we are seeing are part of a natural pattern. One way in which scientists are able to understand what is happening at present, is to understand the earth's *past* atmosphere, and the factors that affected its temperature.

So how, you may be asking, do we know what the earth's *past* climate was like? One method that is used is **ice core drilling**. Antarctica is the coldest continent on earth, and because of this there is very little melting that takes place. Over thousands of years, ice has accumulated in layers and has become more and more compacted as new ice is added. This is partly why Antarctica is also on average one of the *highest* continents! On average, the ice sheet that covers Antarctica is 2500 m thick, and at its deepest location, is 4700 m thick.

As the snow is deposited on top of the ice sheet each year, it traps different chemicals and impurities which are dissolved in the ice. The ice and impurities hold information about the Earth's environment and climate at the time that the ice was deposited. Drilling an ice core from the surface down, is like taking a journey back in time. The deeper into the ice you venture, the older the layer of ice. By analysing the gases and oxygen isotopes that are present (along with many other techniques) in the ice at various points in the earth's history, scientists can start to piece together a picture of what the earth's climate must have been like.

One of the most well known ice cores was the one drilled at a Russian station called **Vostok** in central Antarctica. So far, data has been gathered for dates as far back as 160 000 years!

Activity :: Case Study : Looking at past climatic trends

Make sure that you have read the 'Information box' on ice core drilling before you try this activity.

The values in the table below were extrapolated from data obtained by scientists studying the Vostok ice core. 'Local temperature change' means by how much the temperature at that time was different from what it is today. For example, if the local temperature change 160 000 years ago was -9°C, this means that atmospheric temperatures at that time were 9°C *lower* than what they are today. 'ppm' means 'parts per million' and is a unit of measurement for gas concentrations.

Years before present (x 1000)	Local temperature change (°C)	Carbon dioxide (ppm)
160	-9	190
150	-10	205
140	-10	240
130	-3	280
120	+1	278
110	-4	240
100	-8	225
90	-5	230
80	-6	220
70	-8	250
60	-9	190
50	-7	220
40	-8	180
30	-7	225
20	-9	200
10	-2	260
0 (1850)	-0.5	280
Present		371

Questions

1. On the same set of axes, draw graphs to show how temperature and carbon dioxide concentrations have changed over the last 160 000 years. Hint: 'Years before present' will go on the x-axis, and should be given *negative* values.
2. Compare the graphs that you have drawn. What do you notice?
3. Is there a relationship between temperature and the atmospheric concentration of carbon dioxide?
4. Do these graphs *prove* that temperature changes are determined by the concentration of gases such as carbon dioxide in the atmosphere? Explain your answer.
5. What other factors might you need to consider when analysing climatic trends?

22.3.3 The consequences of global warming

Activity :: Group Discussion : The impacts of global warming

In groups of 3-4, read the following extracts and then answer the questions that follow.

By 2050 Warming to Doom Million Species, Study Says

By 2050, rising temperatures exacerbated by human-induced belches of carbon dioxide and other greenhouse gases could send more than a million of Earth's land-dwelling plants and animals down the road to extinction, according to a recent study. "Climate change now represents at least as great a threat to the number of species surviving on Earth as habitat-destruction and modification," said Chris Thomas, a conservation biologist at the University of Leeds in the United Kingdom.

The researchers worked independently in six biodiversity-rich regions around the world, from Australia to South Africa, plugging field data on species distribution and regional climate into computer models that simulated the

ways species' ranges are expected to move in response to temperature and climate changes. According to the researchers' collective results, the predicted range of climate change by 2050 will place 15 to 35 percent of the 1 103 species studied at risk of extinction.

National Geographic News, 12 July 2004

Global Warming May Dry Up Africa's Rivers, Study Suggests

Many climate scientists already predict that less rain will fall annually in parts of Africa within 50 years due to global warming. Now new research suggests that even a small decrease in rainfall on the continent could cause a drastic reduction in river water, the lifeblood for rural populations in Africa.

A decrease in water availability could occur across about 25 percent of the continent, according to the new study. Hardest hit would be areas in northwestern and southern Africa, with some of the most serious effects striking large areas of Botswana and South Africa.

To predict future rainfall, the scientists compared 21 of what they consider to be the best climate change models developed by research teams around the world. On average, the models forecast a 10 to 20% drop in rainfall in northwestern and southern Africa by 2070. With a 20% decrease, Cape Town would be left with just 42% of its river water, and "Botswana would completely dry up," de Wit said. In parts of northern Africa, river water levels would drop below 50%.

Less river water would have serious implications not just for people but for the many animal species whose habitats rely on regular water supplies.

National Geographic News, 3 March 2006

Discussion questions

1. What is meant by 'biodiversity'?
2. Explain why global warming is likely to cause a *loss of biodiversity*.
3. Why do you think a loss of biodiversity is of such concern to conservationists?
4. Suggest some plant or animal species in South Africa that you think might be particularly vulnerable to extinction if temperatures were to rise significantly. Explain why you chose these species.
5. In what way do people, animals and plants rely on river water?
6. What effect do you think a 50% drop in river water level in some parts of Africa would have on the people living in these countries?
7. Discuss some of the other likely impacts of global warming that we can expect (e.g. sea level rise, melting of polar ice caps, changes in ocean currents).

22.3.4 Taking action to combat global warming

Global warming is a major concern at present. A number of organisations, panels and research bodies have been working to gather accurate and relevant information so that a true picture of our current situation can be painted. One important orgaisation that you may have heard of is the **Intergovernmental Panel on Climate Change** (IPCC). The IPCC was established in 1988 by two United Nations organizations, the World Meteorological Organization (WMO) and the United Nations Environment Programme (UNEP), to evaluate the risk of climate change brought on by humans. You may also have heard of the **Kyoto Protocol**, which will be discussed a little later.

Activity :: Group Discussion : World carbon dioxide emissions

The data in the table below shows carbon dioxide emissions from the consumption of fossil fuels (in million metric tons of carbon dioxide).

Region or Country	1980	1985	1990	1995	2000	2004
United States	4754	4585	5013	5292	5815	5912
Brazil	186	187	222	288	345	336
France	487	394	368	372	399	405
UK	608	588	598	555	551	579
Saudi Arabia	175	179	207	233	288	365
Botswana	1.26	1.45	2.68	3.44	4.16	3.83
South Africa	234	298	295	344	378	429
India	299	439	588	867	1000	1112
World Total	18333	19412	21426	22033	23851	27043

Questions

1. Using a coloured pen, highlight those countries that are 'developed' and those that are 'developing'.
2. Explain why CO_2 emissions are so much higher in developed countries than in developing countries.
3. How does South Africa compare to the other developing countries, and also to the developed countries?

Carbon dioxide emissions are a major problem worldwide. The **Kyoto Protocol** was signed in Kyoto, Japan in December 1997. Its main objective was to reduce global greenhouse gas emissions by encouraging countries to become signatories to the guidelines that had been laid out in the protocol. These guidelines set targets for the world's major producers to reduce their emissions within a certain time. However, some of the worst contributors to greenhouse gas emissions (e.g. USA) were not prepared to sign the protocol, partly because of the potential effect this would have on the country's economy, which relies on industry and other 'high emission' activities.

Panel discussion

Form groups with 5 people in each. Each person in the group must adopt one of the following roles during the discussion:

- the owner of a large industry
- an environmental scientist
- an economist
- a politician
- a chairperson for the discussion

In your group, you are going to discuss some of the economic and environmental implications for a country that decides to sign the Kyoto Protocol. Each person will have the opportunity to express the view of the character they have adopted. You may ask questions of the other people, or challenge their ideas, provided that you ask permission from the chairperson first.

22.4 Summary

- The **atmosphere** is the layer of gases that surrounds Earth. These gases are important in sustaining life, regulating temperature and protecting Earth from harmful radiation.
- The gases that make up the atmosphere are nitrogen, oxygen, carbon dioxide and others e.g. water vapour, methane.
- There are four layer in the atmosphere, each with their own characteristics.

- The **troposphere** is the lowest layer and here, temperature decreases with an increase in altitude. The troposphere is where weather occurs.

- The next layer is the **stratosphere** where temperature increases with an increase in altitude because of the presence of ozone in this layer, and the direct heating from the sun.

- The depletion of the ozone layer is largely because of CFC's, which break down ozone through a series of chemical reactions.

- The **mesosphere** is characterised by very cold temperatures and meteor collisions. The mesosphere contains high concentrations of metal atoms.

- In the **thermosphere**, neutral atoms are ionised by UV and X-ray radiation from the sun. Temperature increases with an increase in altitude because of the energy that is released during this ionisation process, which occurs mostly in the upper thermosphere.

- The thermosphere is also known as the **ionosphere**, and is the part of the atmosphere where radio waves can be transmitted.

- The **auroras** are bright coloured skies that occur when charged particles collide with atoms in the upper atmosphere. Depending on the type of atom, energy is released as light at different wavelengths.

- The Earth is heated by radiation from the sun. Incoming radiation has a short wavelength and some is absorbed directly by the Earth's surface. However, a large amount of energy is re-radiated as longwave infrared radiation.

- **Greenhouse gases** such as carbon dioxide, water vapour and methane absorb infrared radiation and re-emit it back towards the Earth's surface. In this way, the bottom layers of the atmsophere are kept much warmer than they would be if all the infrared radiation was lost.

- Human activities such as the burning of fossil fuels, increase the concentration of greenhouse gases in the atmosphere and may contribute towards **global warming**.

- Some of the impacts of global warming include changing climate patterns, rising sea levels and a loss of biodiversity, to name a few. Interventions are needed to reduce this phenomenon.

?

Exercise: Summary Exercise

1. The atmosphere is a relatively thin layer of gases which support life and provide protection to living organisms. The force of gravity holds the atmosphere against the earth. The diagram below shows the temperatures associated with the various layers that make up the atmosphere and the altitude (height) from the earth's surface.

(a) Write down the names of the layers A, B and D of the atmosphere.

(b) In which one of the layers of the atmosphere is ozone found?

(c) Give an explanation for the decrease in temperature as altitude increases in layer A.

(d) In layer B, there is a steady increase in temperature as the altitude increases. Write down an explanation for this trend.

2. **Planet Earth in Danger**

It is now accepted that greenhouse gases are to blame for planet earth getting warmer. The increase in the number of sudden floods in Asia and droughts in Africa; the rising sea level and increasing average temperatures are global concerns. Without natural greenhouse gases,like carbon dioxide and water vapour,life on earth is not possible. However, the increase in levels of carbon dioxide in the atmosphere since the Industrial Revolution is of great concern. Greater disasters are to come, which will create millions of climate refugees. It is our duty to take action for the sake of future generations who will pay dearly for the wait-and-see attitude of the current generation. Urgent action to reduce waste is needed. Global warming is a global challenge and calls for a global response now, not later.

(Adapted from a speech by the French President, Jacques Chirac)

(a) How do greenhouse gases, such as carbon dioxide, heat up the earth's surface?

(b) Draw a Lewis structure for the carbon dioxide molecule

(c) The chemical bonds within the carbon dioxide molecule are polar. Support this statement by performing a calculation using the table of electronegativities.

(d) Classify the carbon dioxide molecule as polar or non-polar. Give a reason for your answer.

(e) Suggest ONE way in which YOU can help to reduce the emissions of greenhouse gases.

3. Plants need carbon dioxide (CO_2) to manufacture food. However, the engines of motor vehicles cause too much carbon dioxide to be released into the atmosphere.

(a) State the possible consequence of having too much carbon dioxide in the atmosphere.

(b) Explain **two** possible effects on humans if the amount of carbon dioxide in the atmosphere becomes too low.

(DoE Exemplar Paper Grade 11, 2007)

Chapter 23

The Chemical Industry - Grade 12

23.1 Introduction

The chemical industry has been around for a very long time, but not always in the way we think of it today! Dyes, perfumes, medicines and soaps are all examples of products that have been made from chemicals that are found in either plants or animals. However, it was not until the time of the Industrial Revolution that the chemical industry as we know it today began to develop. At the time of the Industrial Revolution, the human population began to grow very quickly and more and more people moved into the cities to live. With this came an increase in the need for things like paper, glass, textiles and soaps. On the farms, there was a greater demand for fertilisers to help produce enough food to feed all the people in cities and rural areas. Chemists and engineers responded to these growing needs by using their technology to produce a variety of new chemicals. This was the start of the chemical industry.

In South Africa, the key event that led to the growth of the chemical industry was the discovery of diamonds and gold in the late 1800's. Mines needed explosives so that they could reach the diamonds and gold-bearing rock, and many of the main chemical companies in South Africa developed to meet this need for explosives. In this chapter, we are going to take a closer look at one of South Africa's major chemical companies, **Sasol**, and will also explore the **chloralkali** and **fertiliser** industries.

23.2 Sasol

Oil and natural gas are important fuel resources. Unfortunately, South Africa has no large oil reserves and, until recently, had very little natural gas. One thing South Africa *does* have however, is large supplies of **coal**. Much of South Africa's chemical industry has developed because of the need to produce oil and gas from coal, and this is where Sasol has played a very important role.

Sasol was established in 1950, with its main aim being to convert low grade coal into petroleum (crude oil) products and other chemical feedstocks. A 'feedstock' is something that is used to make another product. Sasol began producing oil from coal in 1955.

Interesting Fact

The first interest in coal chemistry started as early as the 1920's. In the early 1930's a research engineer called Etienne Rousseau was employed to see whether oil could be made from coal using a new German technology called the **Fischer-Tropsch** process. After a long time, and after many negotiations, Rousseau was given the rights to operate a plant using this new process. As a result, the government-sponsored 'South African Coal, Oil and Gas Corporation Ltd'

(commonly called 'Sasol') was formed in 1950 to begin making oil from coal. A manufacturing plant was established in the Free State and the town of **Sasolburg** developed around this plant. Production began in 1955. In 1969, the **Natref** crude oil refinery was established, and by 1980 and 1982 Sasol Two and Sasol Three had been built at **Secunda**.

23.2.1 Sasol today: Technology and production

Today, Sasol is an oil and gas company with diverse chemical interests. Sasol has three main areas of operation: Firstly, **coal to liquid fuels technology**, secondly the production of **crude oil** and thirdly the conversion of **natural gas to liquid fuel**.

1. **Coal to liquid fuels**

   Sasol is involved in mining coal and converting it into synthetic fuels, using the **Fischer-Tropsch** technology. Figure 23.1 is a simplified diagram of the process that is involved.

Figure 23.1: The gasification of coal to produce liquid fuels

Coal gasification is also known as the **Sasol/Lurgi** gasification process, and involves converting low grade coal to a synthesis gas. Low grade coal has a low percentage carbon, and contains other impurities. The coal is put under extremely high pressure and temperature in the presence of steam and oxygen. The gas that is produced has a high concentration of hydrogen (H_2) and carbon monoxide (CO). That is why it is called a 'synthesis gas', because it is a mixture of more than one gas.

In the **Sasol Advanced Synthol (SAS) reactors**, the gas undergoes a high temperature Fischer-Tropsch conversion. Hydrogen and carbon monoxide react under high pressure and temperature and in the presence of an iron catalyst, to produce a range of hydrocarbon products. Below is the generalised equation for the process. Don't worry too much about the numbers that you see in front of the reactants and products. It is enough just to see that the reaction of hydrogen and carbon monoxide (the two gases in the *synthesis gas*) produces a hydrocarbon and water.

$$(2n+1)H_2 + nCO \rightarrow C_nH_{2n+2} + nH_2O$$

A range of hydrocarbons are produced, including petrol, diesel, jet fuel, propane, butane, ethylene, polypropylene, alcohols and acetic acids.

Important: Different types of fuels

It is important to understand the difference between types of fuels and the terminology that is used for them. The table below summarises some of the fuels that will be mentioned in this chapter.

Compound	Description
Petroleum (crude oil)	A naturally occurring liquid that forms in the earth's lithosphere (see section 21.9 in chapter 21). It is a mixture of hydrocarbons, mostly alkanes, ranging from C_5H_{12} to $C_{18}H_{38}$.
Natural gas	Natural gas has the same origin as petroleum, but is made up of shorter hydrocarbon chains.
Paraffin wax	This is made up of longer hydrocarbon chains, making it a solid compound.
Petrol (gasoline)	A liquid fuel that is derived from petroleum, but which contains extra additives to increase the octane rating of the fuel. Petrol is used as a fuel in combustion engines.
Diesel	Diesel is also derived from petroleum, but is used in diesel engines.
Liquid Petroleum Gas (LPG)	LPG is a mixture of hydrocarbon gases, and is used as a fuel in heating appliances and vehicles. Some LPG mixtures contain mostly propane, while others are mostly butane. LPG is manufactured when crude oil is refined, or is extracted from natural gas supplies in the ground.
Paraffin	This is a technical name for the alkanes, but refers specifically to the *linear* alkanes. *Isoparaffin* refers to non-linear alkanes.
Jet fuel	A type of aviation fuel designed for use in jet engined aircraft. It is an oil-based fuel and contains additives such as antioxidants, corrosion inhibitors and icing inhibitors.

You will notice in the diagram that Sasol doesn't only produce liquid fuels, but also a variety of other chemical products. Sometimes it is the synthetic fuels themselves that are used as feedstocks to produce these chemical products. This is done through processes such as **hydrocracking** and **steamcracking**. Cracking is when heavy hydrocarbons are converted to simpler light hydrocarbons (e.g. LPG and petrol) through the breaking of C-C bonds. A heavy hydrocarbon is one that has a high number of hydrogen and carbon atoms (more solid), and a light hydrocarbon has fewer hydrogen and carbon atoms and is either a liquid or a gas.

Definition: Hydrocracking
Hydrocracking is a cracking process that is assisted by the presence of an elevated partial pressure of hydrogen gas. It produces chemical products such as ethane, LPG, isoparaffins, jet fuel and diesel.

Definition: Steam cracking
Steam cracking occurs under very high temperatures. During the process, a liquid or gaseous hydrocarbon is diluted with steam and then briefly heated in a furnace at a temperature of about 850^0C. Steam cracking is used to convert *ethane* to *ethylene*. Ethylene is a chemical that is needed to make plastics. Steam cracking is also used to make propylene, which is an important fuel gas.

2. **Production of crude oil**

   Sasol obtains crude oil off the coast of Gabon (a country in West Africa) and refines this at the Natref refinery (figure 23.2). Sasol also sells liquid fuels through a number of service stations.

Figure 23.2: Crude oil is refined at Sasol's Natref refinery to produce liquid fuels

3. **Liquid fuels from natural gas**

   Sasol produces natural gas in Mozambique and is expanding its 'gas to fuel' technology. The gas undergoes a complex process to produce linear-chained hydrocarbons such as waxes and paraffins (figure 23.3).

Figure 23.3: Conversion of natural gas to liquid fuels

In the **autothermal reactor**, methane from natural gas reacts with steam and oxygen over an iron-based catalyst to produce a *synthesis gas*. This is a similar process to that involved in coal gasification. The oxygen is produced through the **fractional distillation of air**.

Definition: Fractional distillation
Fractional distillation is the separation of a mixture into its component parts, or fractions. Since air is made up of a number of gases (with the major component being nitrogen), fractional distillation can be used to separate it into these different parts.

The syngas is then passes through a **Sasol Slurry Phase Distillate (SSPD)** process. In this process, the gas is reacted at far lower temperatures than in the SAS reactors. Apart

from hard wax and candle wax, high quality diesel can also be produced in this process. Residual gas from the SSPD process is sold as pipeline gas while some of the lighter hydrocarbons are treated to produce kerosene and paraffin. Ammonia is also produced, which can be used to make fertilisers.

Interesting Fact

Sasol is a major player in the emerging Southern African natural gas industry, after investing 1.2 billion US dollars to develop onshore gas fields in central Mozambique. Sasol has been supplying natural gas from Mozambique's Temane field to customers in South Africa since 2004.

?

Exercise: Sasol processes

Refer to the diagrams summarising the three main Sasol processes, and use these to answer the following questions:

1. Explain what is meant by each of the following terms:
   (a) crude oil
   (b) hydrocarbon
   (c) coal gasification
   (d) synthetic fuel
   (e) chemical feedstock
2. (a) What is diesel?
   (b) Describe two ways in which diesel can be produced.
3. Describe one way in which lighter chemical products such as ethylene, can be produced.
4. Coal and oil play an important role in Sasol's technology.
   (a) In the table below, summarise the similarities and differences between coal, oil and natural gas in terms of how they are formed ('origin'), their general chemical formula and whether they are solid, liquid or gas.

	Coal	Oil	Natural gas
Origin			
General chemical formula			
Solid, liquid or gas			

   (b) In your own words, describe how coal is converted into liquid fuels.
   (c) Explain why Sasol's 'coal to liquid fuels' technology is so important in meeting South Africa's fuel needs.
   (d) Low grade coal is used to produce liquid fuels. What is the main use of higher grade coal in South Africa?

Activity :: Case Study : Safety issues and risk assessments

Safety issues are important to consider when dealing with industrial processes. Read the following extract that appeared in the Business report on 6th February 2006, and then discuss the questions that follow.

> Cape Town - Sasol, the petrochemicals group, was likely to face prosecution on 10 charges of culpable homicide after an explosion at its Secunda plant in 2004 in which 10 people died, a Cape Town labour law specialist said on Friday. The specialist, who did not want to be named, was speaking after the inquiry into the explosion was concluded last Tuesday. It was convened by the labour department.
>
> The evidence led at the inquiry showed a failure on the part of the company to conduct a proper risk assessment and that: Sasol failed to identify hazards associated with a high-pressure gas pipeline running through the plant, which had been shut for extensive maintenance work, in the presence of hundreds of people and numerous machines, including cranes, fitters, contractors, and welding and cutting machines. Because there had never been a risk assessment, the hazard of the high-pressure pipeline had never been identified.
>
> Because Sasol had failed to identify the risk, it did not take any measures to warn people about it, mark the line or take precautions. There had also been inadequacy in planning the shutdown work. In the face of a barrage of criticism for the series of explosions that year, Sasol embarked on a comprehensive programme to improve safety at its operations and appointed Du Pont Safety Resources, the US safety consultancy, to benchmark the petrochemical giant's occupational health and safety performance against international best practice.

1. Explain what is meant by a 'risk assessment'.
2. Imagine that you have been asked to conduct a risk assessment of the Sasol/Lurgi gasification process. What information would you need to know in order to do this assessment?
3. In groups, discuss the importance of each of the following in ensuring the safety of workers in the chemical industry:
   - employing experienced Safety, Health and Environment personnel
   - regular training to identify hazards
   - equipment maintenance and routine checks
4. What other precautions would you add to this list to make sure that working conditions are safe?

23.2.2 Sasol and the environment

From its humble beginnings in 1950, Sasol has grown to become a major contributor towards the South African economy. Today, the industry produces more than 150 000 barrels of fuels and petrochemicals per day, and meets more than 40% of South Africa's liquid fuel requirements. In total, more than 200 fuel and chemical products are manufactured at Sasolburg and Secunda, and these products are exported to over 70 countries worldwide. This huge success is largely due to Sasol's ability to diversify its product base. The industry has also helped to provide about 170 000 jobs in South Africa, and contributes around R40 billion to the country's Gross Domestic Product (GDP).

However, despite these obvious benefits, there are always environmental costs associated with industry. Apart from the vast quantities of resources that are needed in order for the industry to operate, the production process itself produces waste products and pollutants.

Exercise: Consumption of resources

Any industry will always use up huge amounts of resources in order to function effectively, and the chemical industry is no exception. In order for an industry to operate, some of the major resources that are needed are **energy** to drive many of the processes, **water**, either as a coolant or as part of a process and **land** for mining or operations.

Refer to the data table below which shows Sasol's water use between 2002 and 2005 (*Sasol Sustainable Development Report 2005*), and answer the questions that follow.

Water use $(1000m^3)$	2002	2003	2004	2005
River water	113 722	124 179	131 309	124 301
Potable water	15 126	10 552	10 176	10 753
Total	**157 617**	**178 439**	**173 319**	**163 203**

1. Explain what is meant by 'potable' water.
2. Describe the trend in Sasol's water use that you see in the above statistics.
3. Suggest possible reasons for this trend.
4. List some of the environmental impacts of using large amounts of river water for industry.
5. Suggest ways in which these impacts could be reduced

Exercise: Industry and the environment

Large amounts of gases and pollutants are released during production, and when the fuels themselves are used. Refer to the table below, which shows greenhouse gas and atmospheric pollution data for Sasol between 2002 and 2005, and then answer the questions that follow. (*Source: Sasol Sustainable Development Report 2005*)

Greenhouse gases and air pollutants (kilotonnes)	2002	2003	2004	2005
Carbon dioxide (CO_2)	57 476	62 873	66 838	60 925
Hydrogen sulfide (H_2S)	118	105	102	89
Nitrogen oxides (NO_x)	168	173	178	166
Sulfur dioxide (SO_2)	283	239	261	222

1. Draw line graphs to show how the quantity of each pollutant produced has changed between 2002 and 2005.
2. Describe what you see in the graphs, and suggest a reason for this trend.
3. Explain what is meant by each of the following terms:
   (a) greenhouse gas
   (b) global warming
4. Describe some of the possible effects of global warming.
5. When sulfur dioxide is present in the atmosphere, it may react with water vapour to produce weak *sulfuric acid*. In the same way, nitrogen dioxide and water vapour react to form *nitric acid*. These reactions in the atmosphere may cause **acid rain**. Outline some of the possible consequences of acid rain.
6. Many industries are major contributors towards environmental problems such as global warming, environmental pollution, over-use of resources and acid rain. Industries are in a difficult position: On one hand they must meet the ever increasing demands of society, and on the other, they must achieve this with as little environmental impact as possible. This is a huge challenge.

- Work in groups of 3-4 to discuss ways in which industries could be encouraged (or in some cases forced) to reduce their environmental impact.
- Elect a spokesperson for each group, who will present your ideas to the class.
- Are the ideas suggested by each group practical?
- How easy or difficult do you think it would be to implement these ideas in South Africa?

Interesting Fact

Sasol is very aware of its responsibility towards creating cleaner fuels. From 1st January 2006, the South African government enforced a law to prevent lead from being added to petrol. Sasol has complied with this. One branch of Sasol, **Sasol Technology** also has a bio-diesel research and development programme focused on developing more environmentally friendly forms of diesel. One way to do this is to use renewable resources such as soybeans to make diesel. Sasol is busy investigating this new technology.

23.3 The Chloralkali Industry

The chlorine-alkali (chloralkali) industry is an important part of the chemical industry, and produces **chlorine** and **sodium hydroxide** through the electrolysis of salt (NaCl). The main raw material is **brine** which is a saturated solution of sodium chloride (NaCl) that is obtained from natural salt deposits.

The products of this industry have a number of important uses. **Chlorine** is used to purify water, and is used as a disinfectant. It is also used in the manufacture of many every-day items such as hypochlorous acid, which is used to kill bacteria in drinking water. Chlorine is also used in paper production, antiseptics, food, insecticides, paints, petroleum products, plastics (such as polyvinyl chloride or PVC), medicines, textiles, solvents, and many other consumer products. Many chemical products such as chloroform and carbon tetrachloride also contain chlorine.

Sodium hydroxide (also known as 'caustic soda') has a number of uses, which include making soap and other cleaning agents, purifying bauxite (the ore of aluminium), making paper and making rayon (artificial silk).

23.3.1 The Industrial Production of Chlorine and Sodium Hydroxide

Chlorine and sodium hydroxide can be produced through a number of different reactions. However, one of the problems is that when chlorine and sodium hydroxide are produced together, the chlorine combines with the sodium hydroxide to form chlorate (ClO^-) and chloride (Cl^-) ions. This produces sodium chlorate, NaClO, a component of household bleach. To overcome this problem the chlorine and sodium hydroxide must be separated from each other so that they don't react. There are three industrial processes that have been designed to overcome this problem, and to produce chlorine and sodium hydroxide. All three methods involve **electrolytic cells** (chapter 17).

Important: Electrolytic cells

Electrolytic cells are used to split up or loosen ions. They are made up of an **electrolyte** and two electrodes, the **cathode** and the **anode**. An electrolytic cell is activated by applying an external electrical current. This creates an electrical potential across the cathode and anode, and forces a chemical reaction to take place in the electrolyte. Cations flow towards the cathode and are reduced. Anions flow to the anode and are oxidised. Two new products are formed, one product at the cathode and one at the anode.

1. **The Mercury Cell**

   In the mercury-cell (figure 23.4), brine passes through a chamber which has a carbon electrode (the anode) suspended from the top. Mercury flows along the floor of this chamber and acts as the cathode. When an electric current is applied to the circuit, chloride ions in the electrolyte are oxidised to form chlorine gas.

$$2\mathrm{Cl}^-_{(\mathrm{aq})} \rightarrow \mathrm{Cl}_{2(\mathrm{g})} + 2\mathrm{e}^-$$

   At the cathode, sodium ions are reduced to sodium.

$$2\mathrm{Na}^+_{(\mathrm{aq})} + 2\mathrm{e}^- \rightarrow 2\mathrm{Na}_{(\mathrm{Hg})}$$

   The sodium dissolves in the mercury, forming an amalgam of sodium and mercury. The amalgam is then poured into a separate vessel, where it decomposes into sodium and mercury. The sodium reacts with water in the vessel and produces sodium hydroxide (caustic soda) and hydrogen gas, while the mercury returns to the electrolytic cell to be used again.

$$2\mathrm{Na}_{(\mathrm{Hg})} + 2\mathrm{H_2O}_{(\mathrm{l})} \rightarrow 2\mathrm{NaOH}_{(\mathrm{aq})} + \mathrm{H}_{2(\mathrm{g})}$$

Figure 23.4: The Mercury Cell

This method, however, only produces a fraction of the chlorine and sodium hydroxide that is used by industry as it has certain disadvantages: mercury is expensive and toxic, and although it is returned to the electrolytic cell, some always escapes with the brine that has been used. The mercury reacts with the brine to form mercury(II) chloride. In the past this effluent was released into lakes and rivers, causing mercury to accumulate in fish and other animals feeding on the fish. Today, the brine is treated before it is discharged so that the environmental impact is lower.

2. **The Diaphragm Cell**

   In the diaphragm-cell (figure 23.5), a porous diaphragm divides the electrolytic cell, which contains brine, into an anode compartment and a cathode compartment. The brine is introduced into the anode compartment and flows through the diaphragm into the cathode compartment. When an electric current passes through the brine, the salt's chlorine ions and sodium ions move to the electrodes. Chlorine gas is produced at the anode. At the cathode, sodium ions react with water, forming caustic soda and hydrogen gas. Some salt remains in the solution with the caustic soda and can be removed at a later stage.

Figure 23.5: Diaphragm Cell

   This method uses less energy than the mercury cell, but the sodium hydroxide is not as easily concentrated and precipitated into a useful substance.

Interesting Fact

To separate the chlorine from the sodium hydroxide, the two half-cells were traditionally separated by a porous asbestos diaphragm, which needed to be replaced every two months. This was damaging to the environment, as large quantities of asbestos had to be disposed. Today, the asbestos is being replaced by other polymers which do not need to be replaced as often.

3. **The Membrane Cell**

   The membrane cell (figure 23.6) is very similar to the diaphragm cell, and the same reactions occur. The main difference is that the two electrodes are separated by an ion-selective membrane, rather than by a diaphragm. The structure of the membrane is such that it allows cations to pass through it between compartments of the cell. It does not allow anions to pass through. This has nothing to do with the size of the pores, but rather with the charge on the ions. Brine is pumped into the anode compartment, and only the positively charged sodium ions pass into the cathode compartment, which contains pure water.

   At the positively charged anode, Cl^- ions from the brine are oxidised to Cl_2 gas.

$$2\mathrm{Cl}^- \rightarrow \mathrm{Cl}_{2(g)} + 2\mathrm{e}^-$$

   At the negatively charged cathode, hydrogen ions in the water are reduced to hydrogen gas.

Figure 23.6: Membrane Cell

$$2\mathrm{H}^{+}_{(\mathrm{aq})} + 2\mathrm{e}^{-} \rightarrow \mathrm{H}_{2(\mathrm{g})}$$

The Na^+ ions flow through the membrane to the cathode compartment and react with the remaining hydroxide (OH^-) ions from the water to form sodium hydroxide (NaOH). The chloride ions cannot pass through, so the chlorine does not come into contact with the sodium hydroxide in the cathode compartment. The sodium hydroxide is removed from the cell. The overall equation is as follows:

$$2\mathrm{NaCl} + 2\mathrm{H}_2\mathrm{0} \rightarrow \mathrm{Cl}_2 + \mathrm{H}_2 + 2\mathrm{NaOH}$$

The advantage of using this method is that the sodium hydroxide that is produced is very pure because it is kept separate from the sodium chloride solution. The caustic soda therefore has very little salt contamination. The process also uses less electricity and is cheaper to operate.

?

Exercise: The Chloralkali industry

1. Refer to the flow diagram below which shows the reactions that take place in the membrane cell, and then answer the questions that follow.

(a) What liquid is present in the cathode compartment at (a)?
(b) Identify the gas that is produced at (b).
(c) Explain one feature of this cell that allows the Na^+ and OH^- ions to react at (c).
(d) Give a balanced equation for the reaction that takes place at (c).

2. Summarise what you have learnt about the three types of cells in the chloralkali industry by completing the table below:

	Mercury cell	**Diaphragm cell**	**Membrane cell**
Main raw material			
Mechanism of separating Cl_2 and NaOH			
Anode reaction			
Cathode reaction			
Purity of NaOH produced			
Energy consumption			
Environmental impact			

23.3.2 Soaps and Detergents

Another important part of the chloralkali industry is the production of **soaps** and **detergents**. You will remember from an earlier chapter, that water has the property of *surface tension*. This means that it tends to bead up on surfaces and this slows down the wetting process and makes cleaning difficult. You can observe this property of surface tension when a drop of water falls onto a table surface. The drop holds its shape and does not spread. When cleaning, this surface tension must be reduced so that the water can spread. Chemicals that are able to do this

are called **surfactants**. Surfactants also loosen, disperse and hold particles in suspension, all of which are an important part of the cleaning process. Soap is an example of one of these surfactants. Detergents contain one or more surfactants. We will go on to look at these in more detail.

Definition: Surfactant
A surfactant is a wetting agent that lowers the surface tension of a liquid, allowing it to spread more easily.

1. **Soaps**

   In chapter 10, a number of important biological macromolecules were discussed, including carbohydrates, proteins and nucleic acids. **Fats** are also biological macromolecules. A fat is made up of an alcohol called glycerol, attached to three fatty acids (figure 23.7). Each **fatty acid** is made up of a carboxylic acid attached to a long hydrocarbon chain. An **oil** has the same structure as a fat, but is a liquid rather than a solid. Oils are found in plants (e.g. olive oil, sunflower oil) and fats are found in animals.

Figure 23.7: The structure of a fat, composed of an alcohol and three fatty acids

To make soap, sodium hydroxide (NaOH) or potassium hydroxide (KOH) must be added to a fat or an oil. During this reaction, the glycerol is separated from the hydrocarbon chain in the fat, and is replaced by either potassium or sodium ions (figure 23.8). Soaps are the water-soluble sodium or potassium salts of fatty acids.

Interesting Fact

Soaps can be made from either fats or oils. Beef fat is a common source of fat, and vegetable oils such as palm oil are also commonly used.

Fatty acids consist of two parts: a carboxylic acid group and a hydrocarbon chain. The hydrocarbon chain is *hydrophobic*, meaning that it is repelled by water. However, it is attracted to grease, oils and other dirt. The carboxylic acid is *hydrophilic*, meaning that it is attracted to water. Let's imagine that we have added soap to water in order to clean a dirty rugby jersey. The hydrocarbon chain will attach itself to the soil particles in the

Figure 23.8: Sodium hydroxide reacts with a fat to produce glycerol and sodium salts of the fatty acids

jersey, while the carboxylic acid will be attracted to the water. In this way, the soil is pulled free of the jersey and is suspended in the water. In a washing machine or with vigourous handwashing, this suspension can be rinsed off with clean water.

Definition: Soap
Soap is a surfactant that is used with water for washing and cleaning. Soap is made by reacting a fat with either sodium hydroxide (NaOH) or potassium hydroxide (KOH).

2. **Detergents**

Definition: Detergent
Detergents are compounds or mixtures of compounds that are used to assist cleaning. The term is often used to distinguish between soap and other chemical surfactants for cleaning.

Detergents are also cleaning products, but are composed of one or more surfactants. Depending on the type of cleaning that is needed, detergents may contain one or more of the following:

- *Abrasives* to scour a surface.
- *Oxidants* for bleaching and disinfection.
- *Enzymes* to digest proteins, fats or carbohydrates in stains. These are called *biological detergents*.

Exercise: The choralkali industry

1. The diagram above shows the sequence of steps that take place in the mercury cell.
   (a) Name the 'raw material' in step 1.
   (b) Give the chemical equation for the reaction that produces chlorine in step 2.
   (c) What other product is formed in step 2.
   (d) Name the reactants in step 4.
2. Approximately 30 million tonnes of chlorine are used throughout the world annually. Chlorine is produced industrially by the electrolysis of brine. The diagram represents a membrane cell used in the production of Cl_2 gas.

   (a) What ions are present in the electrolyte in the left hand compartment of the cell?
   (b) Give the equation for the reaction that takes place at the anode.
   (c) Give the equation for the reaction that takes place at the cathode.
   (d) What ion passes through the membrane while these reactions are taking place?
   Chlorine is used to purify drinking water and swimming pool water. The substance responsible for this process is the weak acid, hypochlorous acid (HOCl).
   (e) One way of putting HOCl into a pool is to bubble chlorine gas through the water. Give an equation showing how bubbling $Cl_2(g)$ through water produces HOCl.
   (f) A common way of treating pool water is by adding 'granular chlorine'. Granular chlorine consists of the salt calcium hypochlorite, $Ca(OCl)_2$. Give an equation showing how this salt dissolves in water. Indicate the phase of each substance in the equation.
   (g) The OCl^- ion undergoes hydrolysis , as shown by the following equation:
   $$OCl^- + H_2O \Leftrightarrow HOCl + OH^-$$
   Will the addition of granular chlorine to pure water make the water acidic, basic or will it remain neutral? Briefly explain your answer.

   (IEB Paper 2, 2003)

23.4 The Fertiliser Industry

23.4.1 The value of nutrients

Nutrients are very important for life to exist. An **essential nutrient** is any chemical element that is needed for a plant to be able to grow from a seed and complete its life cycle. The same is true for animals. A *macronutrient* is one that is required in large quantities by the plant or animal, while a *micronutrient* is one that only needs to be present in small amounts for a plant or an animal to function properly.

Definition: Nutrient
A nutrient is a substance that is used in an organism's metabolism or physiology and which must be taken in from the environment.

In plants, the macronutrients include carbon (C), hydrogen (H), oxygen (O), nitrogen (N), phosphorus (P) and potassium (K). The source of each of these nutrients for plants, and their function, is summarised in table 23.1. Examples of micronutrients in plants include iron, chlorine, copper and zinc.

Table 23.1: The source and function of the macronutrients in plants

Nutrient	Source	Function
Carbon	Carbon dioxide in the air	Component of organic molecules such as carbohydrates, lipids and proteins
Hydrogen	Water from the soil	Component of organic molecules
Oxygen	Water from the soil	Component of organic molecules
Nitrogen	Nitrogen compounds in the soil	Part of plant proteins and chlorophyll. Also boosts plant growth.
Phosphorus	Phosphates in the soil	Needed for photosynthesis, blooming and root growth
Potassium	Soil	Building proteins, part of chlorophyll and reduces diseases in plants

Animals need similar nutrients in order to survive. However since animals can't photosynthesise, they rely on plants to supply them with the nutrients they need. Think for example of the human diet. We can't make our own food and so we either need to eat vegetables, fruits and seeds (all of which are direct plant products) or the meat of other animals which would have fed on plants during their life. So most of the nutrients that animals need are obtained either directly or indirectly from plants. Table 23.2 summarises the functions of some of the macronutrients in animals.

Micronutrients also play an important function in animals. Iron for example, is found in haemoglobin, the blood pigment that is responsible for transporting oxygen to all the cells in the body.

Nutrients then, are essential for the survival of life. Importantly, obtaining nutrients starts with plants, which are able either to photosynthesise or to absorb the required nutrients from the soil. It is important therefore that plants are always able to access the nutrients that they need so that they will grow and provide food for other forms of life.

23.4.2 The Role of fertilisers

Plants are only able to absorb soil nutrients in a particular form. Nitrogen for example, is absorbed as **nitrates**, while phosphorus is absorbed as **phosphates**. The **nitrogen cycle** (chapter

Table 23.2: The functions of animal macronutrients

Nutrient	Function
Carbon	Component of organic compounds
Hydrogen	Component of organic compounds
Oxygen	Component of organic compounds
Nitrogen	Component of nucleic acids and proteins
Phosphorus	Component of nucleic acids and phospholipids
Potassium	Helps in coordination and regulating the water balance in the body

19) describes the process that is involved in converting atmospheric nitrogen into a form that can be used by plants.

However, all these natural processes of maintaining soil nutrients take a long time. As populations grow and the demand for food increases, there is more and more strain on the land to produce food. Often, cultivation practices don't give the soil enough time to recover and to replace the nutrients that have been lost. Today, **fertilisers** play a very important role in restoring soil nutrients so that crop yields stay high. Some of these fertilisers are **organic** (e.g. compost, manure and fishmeal), which means that they started off as part of something living. Compost for example is often made up of things like vegetable peels and other organic remains that have been thrown away. Others are **inorganic** and can be made industrially. The advantage of these commercial fertilisers is that the nutrients are in a form that can be absorbed immediately by the plant.

Definition: Fertiliser
A fertiliser is a compound that is given to a plant to promote growth. Fertilisers usually provide the three major plant nutrients and most are applied via the soil so that the nutrients are absorbed by plants through their roots.

When you buy fertilisers from the shop, you will see three numbers on the back of the packet e.g. 18-24-6. These numbers are called the **NPK ratio**, and they give the percentage of nitrogen, phosphorus and potassium in that fertiliser. Depending on the types of plants you are growing, and the way in which you would like them to grow, you may need to use a fertiliser with a slightly different ratio. If you want to encourage root growth in your plant for example, you might choose a fertiliser with a greater amount of phosphorus. Look at the table below, which gives an idea of the amounts of nitrogen, phosphorus and potassium there are in different types of fertilisers. Fertilisers also provide other nutrients such as calcium, sulfur and magnesium.

Table 23.3: Common grades of some fertiliser materials

Description	Grade (NPK %)
Ammonium nitrate	34-0-0
Urea	46-0-0
Bone Meal	4-21-1
Seaweed	1-1-5
Starter fertilisers	18-24-6
Equal NPK fertilisers	12-12-12
High N, low P and medium K fertilisers	25-5-15

23.4.3 The Industrial Production of Fertilisers

The industrial production of fertilisers may involve several processes.

1. **Nitrogen fertilisers**

   Making **nitrogen fertilisers** involves producing *ammonia*, which is then reacted with *oxygen* to produce *nitric acid*. Nitric acid is used to acidify phosphate rock to produce nitrogen fertilisers. The flow diagram below illustrates the processes that are involved. Each of these steps will be examined in more detail.

Figure 23.9: Flow diagram showing steps in the production of nitrogen fertilisers

(a) **The Haber Process**

The Haber process involves the reaction of nitrogen and hydrogen to produce ammonia. **Nitrogen** is produced through the **fractional distillation** of air. Fractional distillation is the separation of a mixture (remember that air is a mixture of different gases) into its component parts through various methods. **Hydrogen** can be produced through **steam reforming**. In this process, a hydrocarbon such as methane reacts with water to form carbon monoxide and hydrogen according to the following equation:

$$\mathrm{CH_4 + H_2O \rightarrow CO + 3H_2}$$

Nitrogen and hydrogen are then used in the Haber process. The equation for the Haber process is:

$$\mathrm{N_2(g) + 3H_2(g) \rightarrow 2NH_3(g)}$$

(The reaction takes place in the presence of an iron (Fe) catalyst under conditions of 200 atmospheres (atm) and 450-500 degrees Celsius)

Interesting Fact

The Haber process developed in the early 20th century, before the start of World War 1. Before this, other sources of nitrogen for fertilisers had included saltpeter ($NaNO_3$) from Chile and guano. Guano is the droppings of seabirds, bats and seals. By the 20th century, a number of methods had been developed to 'fix' atmospheric nitrogen. One of these was the Haber process, and it advanced through the work of two German men, Fritz Haber and Karl Bosch (The process is sometimes also referred to as the 'Haber-Bosch process'). They worked out what the best conditions were in order to get a high yield of ammonia, and found these to be high temperature and high pressure. They also experimented with different catalysts to see which worked best in that reaction. During World War 1, the ammonia that was produced through the Haber process was used to make explosives. One of the advantages for Germany was that, having perfected the Haber process, they did not need to rely on other countries for the chemicals that they needed to make them.

(b) **The Ostwald Process**

The Ostwald process is used to produce nitric acid from ammonia. Nitric acid can then be used in reactions that produce fertilisers. Ammonia is converted to nitric acid in two stages. First, it is oxidised by heating with oxygen in the presence of a platinum catalyst to form nitric oxide and water. This step is strongly exothermic, making it a useful heat source.

$$4NH_3(g) + 5O_2(g) \rightarrow 4NO(g) + 6H_2O(g)$$

Stage two, which combines two reaction steps, is carried out in the presence of water. Initially nitric oxide is oxidised again to yield nitrogen dioxide:

$$2NO(g) + O_2(g) \rightarrow 2NO_2(g)$$

This gas is then absorbed by the water to produce nitric acid. Nitric oxide is also a product of this reaction. The nitric oxide (NO) is recycled, and the acid is concentrated to the required strength.

$$3NO_2(g) + H_2O(l) \rightarrow 2HNO_3(aq) + NO(g)$$

(c) **The Nitrophosphate Process**

The nitrophosphate process involves acidifying phosphate rock with nitric acid to produce a mixture of phosphoric acid and calcium nitrate:

$$Ca_3(PO_4)_2 + 6HNO_3 + 12H_2O \rightarrow 2H_3PO_4 + 3Ca(NO_3)_2 + 12H_2O$$

When calcium nitrate and phosphoric acid react with ammonia, a compound fertiliser is produced.

$$Ca(NO_3)_2 + 4H_3PO_4 + 8NH_3 \rightarrow CaHPO_4 + 2NH_4NO_3 + 8(NH_4)2HPO_4$$

If potassium chloride or potassium sulphate is added, the result will be NPK fertiliser.

(d) **Other nitrogen fertilisers**

- Urea ($(NH_2)_2CO$) is a nitrogen-containing chemical product which is produced on a large scale worldwide. Urea has the highest nitrogen content of all solid nitrogeneous fertilisers in common use (46.4%) and is produced by reacting ammonia with carbon dioxide.

  Two reactions are involved in producing urea:

  i. $2NH_3 + CO_2 \rightarrow H_2N - COONH_4$
  ii. $H_2N - COONH_4 \rightarrow (NH_2)_2CO + H_2O$

- Other common fertilisers are ammonium nitrate and ammonium sulphate. Ammonium nitrate is formed by reacting ammonia with nitric acid.

  $$NH_3 + HNO_3 \rightarrow NH_4NO_3$$

  Ammonium sulphate is formed by reacting ammonia with sulphuric acid.

  $$2NH_3 + H_2SO_4 \rightarrow (NH_4)_2SO_4$$

2. **Phosphate fertilisers**

The production of phosphate fertilisers also involves a number of processes. The first is the production of sulfuric acid through the **contact process**. Sulfuric acid is then used in a reaction that produces phosphoric acid. Phosphoric acid can then be reacted with phosphate rock to produce triple superphosphates.

(a) *The production of sulfuric acid*

Sulfuric acid is produced from sulfur, oxygen and water through the contact process. In the first step, sulfur is burned to produce sulfur dioxide.

$$S(s) + O_2(g) \rightarrow SO_2(g)$$

This is then oxidised to sulfur trioxide using oxygen in the presence of a vanadium(V) oxide catalyst.

$$2SO_2 + O_2(g) \rightarrow 2SO_3(g)$$

Finally the sulfur trioxide is treated with water to produce 98-99% sulfuric acid.

$$SO_3(g) + H_2O(l) \rightarrow H_2SO_4(l)$$

(b) *The production of phosphoric acid*

The next step in the production of phosphate fertiliser is the reaction of sulfuric acid with phosphate rock to produce phosphoric acid (H_3PO_4). In this example, the phosphate rock is fluoropatite ($Ca_5F(PO_4)_3$).

$$Ca_5F(PO_4)_3 + 5H_2SO_4 + 10H_2O \rightarrow 5CaSO_42H_2O + HF + 3H_3PO_4$$

(c) *The production of phosphates and superphosphates*

When concentrated phosphoric acid reacts with ground phosphate rock, triple superphosphate is produced.

$$3Ca_3(PO_4)_2CaF_2 + 4H_3PO_4 + 9H_2O \rightarrow 9Ca(H_2PO_4)_2 + CaF_2$$

3. **Potassium**

Potassium is obtained from **potash**, an impure form of potassium carbonate (K_2CO_3). Other potassium salts (e.g. KCl AND K_2O) are also sometimes included in fertilisers.

23.4.4 Fertilisers and the Environment: Eutrophication

Eutrophication is the enrichment of an ecosystem with chemical nutrients, normally by compounds that contain nitrogen or phosphorus. Eutrophication is considered a form of pollution because it promotes plant growth, favoring certain species over others. In aquatic environments, the rapid growth of certain types of plants can disrupt the normal functioning of an ecosystem, causing a variety of problems. Human society is impacted as well because eutrophication can decrease the resource value of rivers, lakes, and estuaries making recreational activities less enjoyable. Health-related problems can also occur if eutrophic conditions interfere with the treatment of drinking water.

Definition: Eutrophication
Eutrophication refers to an increase in chemical nutrients in an ecosystem. These chemical nutrients usually contain nitrogen or phosphorus.

In some cases, eutrophication can be a natural process that occurs very slowly over time. However, it can also be accelerated by certain human activities. Agricultural runoff, when excess fertilisers are washed off fields and into water, and sewage are two of the major causes of eutrophication. There are a number of impacts of eutrophication.

- *A decrease in biodiversity (the number of plant and animal species in an ecosystem)*

  When a system is enriched with nitrogen, plant growth is rapid. When the number of plants increases in an aquatic system, they can block light from reaching deeper. Plants also consume oxygen for respiration, and if the oxygen content of the water decreases too much, this can cause other organisms such as fish to die.

- *Toxicity*

  Sometimes, the plants that flourish during eutrophication can be toxic and may accumulate in the food chain.

Interesting Fact

South Africa's Department of Water Affairs and Forestry has a 'National Eutrophication Monitoring Programme' which was set up to monitor eutrophication in impoundments such as dams, where no monitoring was taking place.

Despite the impacts, there are a number of ways of preventing eutrophication from taking place. **Cleanup measures** can directly remove the excess nutrients such as nitrogen and phosphorus from the water. Creating **buffer zones** near farms, roads and rivers can also help. These act as filters and cause nutrients and sediments to be deposited there instead of in the aquatic system. **Laws** relating to the treatment and discharge of sewage can also help to control eutrophication. A final possible intervention is **nitrogen testing and modeling**. By assessing exactly how much fertiliser is needed by crops and other plants, farmers can make sure that they only apply just enough fertiliser. This means that there is no excess to run off into neighbouring streams during rain. There is also a cost benefit for the farmer.

Activity :: Discussion : Dealing with the consequences of eutrophication

In many cases, the damage from eutrophication is already done. In groups, do the following:

1. List all the possible consequences of eutrophication that you can think of.
2. Suggest ways to solve these problems, that arise because of eutrophication.

Exercise: Chemical industry: Fertilisers

Why we need fertilisers

There is likely to be a gap between food production and demand in several parts of the world by 2020. Demand is influenced by population growth and urbanisation, as well as income levels and changes in dietary preferences.
The facts are as follows:

- There is an increasing world population to feed
- Most soils in the world used for large-scale, intensive production of crops lack the necessary nutrients for the crops

Conclusion: Fertilisers are needed!

The flow diagram below shows the main steps in the industrial preparation of two important solid fertilisers.

1. Write down the balanced chemical equation for the formation of the brown gas.
2. Write down the name of process Y.
3. Write down the chemical formula of liquid E.
4. Write down the chemical formulae of fertilisers C and D respectively.
   The following extract comes from an article on fertilisers:

   A world without food for its people
   A world with an environment poisoned through the actions of man
   Are two contributing factors towards a disaster scenario.

5. Write down THREE ways in which the use of fertilisers poisons the environment.

23.5 Electrochemistry and batteries

You will remember from chapter 17 that a **galvanic** cell (also known as a *voltaic* cell) is a type of electrochemical cell where a chemical reaction produces electrical energy. The **emf** of a galvanic cell is the difference in voltage between the two half cells that make it up. Galvanic cells have a number of applications, but one of the most important is their use in **batteries**. You will know from your own experience that we use batteries in a number of ways, including cars, torches, sound systems and cellphones to name just a few.

23.5.1 How batteries work

A battery is a device in which **chemical energy** is directly converted to **electrical energy**. It consists of one or more voltaic cells, each of which is made up of two half cells that are connected in series by a conductive electrolyte. The voltaic cells are connected in series in a battery. Each cell has a positive electrode (cathode), and a negative electrode (anode). These do not touch each other but are immersed in a solid or liquid electrolyte.

Each half cell has a net electromotive force (emf) or voltage. The voltage of the battery is the difference between the voltages of the half-cells. This potential difference between the two half cells is what causes an electric current to flow.

Batteries are usually divided into two broad classes:

- *Primary batteries* irreversibly transform chemical energy to electrical energy. Once the supply of reactants has been used up, the battery can't be used any more.
- *Secondary batteries* can be recharged, in other words, their chemical reactions can be reversed if electrical energy is supplied to the cell. Through this process, the cell returns to its original state. Secondary batteries can't be recharged forever because there is a gradual loss of the active materials and electrolyte. Internal corrosion can also take place.

23.5.2 Battery capacity and energy

The **capacity** of a battery, in other words its ability to produce an electric charge, depends on a number of factors. These include:

- **Chemical reactions**

  The chemical reactions that take place in each of a battery's half cells will affect the voltage across the cell, and therefore also its capacity. For example, nickel-cadmium (NiCd) cells measure about 1.2V, and alkaline and carbon-zinc cells both measure about 1.5 volts. However, in other cells such as Lithium cells, the changes in electrochemical potential are much higher because of the reactions of lithium compounds, and so lithium cells can produce as much as 3 volts or more. The concentration of the chemicals that are involved will also affect a battery's capacity. The higher the concentration of the chemicals, the greater the capacity of the battery.

- **Quantity of electrolyte and electrode material in cell**

  The greater the amount of electrolyte in the cell, the greater its capacity. In other words, even if the chemistry in two cells is the same, a larger cell will have a greater capacity than a small one. Also, the greater the surface area of the electrodes, the greater will be the capacity of the cell.

- **Discharge conditions**

  A unit called an **Ampere hour** (Ah) is used to describe how long a battery will last. An ampere hour (more commonly known as an **amp hour**) is the amount of electric charge that is transferred by a current of one ampere for one hour. Battery manufacturers use a standard method to rate their batteries. So, for example, a 100 Ah battery will provide a current of 5 A for a period of 20 hours at room temperature. The capacity of the battery will depend on the rate at which it is discharged or used. If a 100 Ah battery is discharged at 50 A (instead of 5 A), the capacity will be *lower* than expected and the battery will run out *before* the expected 2 hours.

  The relationship between the current, discharge time and capacity of a battery is expressed by **Peukert's law**:

  $$C_p = I^k t$$

  In the equation, 'C_p' represents the battery's capacity (Ah), I is the discharge current (A), k is the Peukert constant and t is the time of discharge (hours).

23.5.3 Lead-acid batteries

In a **lead-acid battery**, each cell consists of electrodes of lead (Pb) and lead (IV) oxide (PbO_2) in an electrolyte of sulfuric acid (H_2SO_4). When the battery discharges, both electrodes turn

into lead (II) sulphate ($PbSO_4$) and the electrolyte loses sulfuric acid to become mostly water.

The chemical half reactions that take place at the anode and cathode when the battery is **discharging** are as follows:

Anode (oxidation): $\mathrm{Pb(s) + SO_4^{2-}(aq) \Leftrightarrow PbSO_4(s) + 2e^-}$ (E^0 = -0.356 V)

Cathode (reduction): $\mathrm{PbO_2(s) + SO_4^{2-}(aq) + 4H^+ + 2e^- \Leftrightarrow PbSO_4(s) + 2H_2O(l)}$ (E^0 = 1.685 V)

The overall reaction is as follows:

$$\mathrm{PbO_2(s) + 4H^+(aq) + 2SO_4^{2-}(aq) + Pb(s) \rightarrow 2PbSO_4(s) + 2H_2O(l)}$$

The emf of the cell is calculated as follows:

EMF = E (cathode)- E (anode)

EMF = +1.685 V - (-0.356 V)

EMF = +2.041 V

Since most batteries consist of six cells, the total voltage of the battery is approximately 12 V.

One of the important things about a lead-acid battery is that it can be **recharged**. The recharge reactions are the *reverse* of those when the battery is discharging.

The lead-acid battery is made up of a number of *plates* that maximise the surface area on which chemical reactions can take place. Each plate is a rectangular grid, with a series of holes in it. The holes are filled with a mixture of lead and sulfuric acid. This paste is pressed into the holes and the plates are then stacked together, with suitable separators between them. They are then placed in the battery container, after which acid is added (figure 23.10).

Figure 23.10: A lead-acid battery

Lead-acid batteries have a number of applications. They can supply high surge currents, are relatively cheap, have a long shelf life and can be recharged. They are ideal for use in cars,

where they provide the high current that is needed by the starter motor. They are also used in forklifts and as standby power sources in telecommunication facilities, generating stations and computer data centres. One of the disadvantages of this type of battery is that the battery's lead must be recycled so that the environment doesn't become contaminated. Also, sometimes when the battery is charging, hydrogen gas is generated at the cathode and this can cause a small explosion if the gas comes into contact with a spark.

23.5.4 The zinc-carbon dry cell

A simplified diagram of a zinc-carbon cell is shown in figure 23.11.

Figure 23.11: A zinc-carbon dry cell

A zinc-carbon cell is made up of an outer zinc container, which acts as the **anode**. The **cathode** is the central carbon rod, surrounded by a mixture of carbon and manganese (IV) oxide (MnO_2). The **electrolyte** is a paste of ammonium chloride (NH_4Cl). A fibrous fabric separates the two electrodes, and a brass pin in the centre of the cell conducts electricity to the outside circuit.

The paste of ammonium chloride reacts according to the following half-reaction:

$$2NH_4^+(aq) + 2e^- \rightarrow 2NH_3(g) + H_2(g)$$

The manganese(IV) oxide in the cell removes the hydrogen produced above, according to the following reaction:

$$2MnO_2(s) + H_2(g) \rightarrow Mn_2O_3(s) + H_2O(l)$$

The combined result of these two reactions can be represented by the following half reaction, which takes place at the cathode:

Cathode: $2NH_4^+(aq) + 2MnO_2(s) + 2e^- \rightarrow Mn_2O_3(s) + 2NH_3(g) + H_2O(l)$

The anode half reaction is as follows:

Anode: $Zn(s) \rightarrow Zn^{2+} + 2e^-$

The overall equation for the cell is:

$$Zn(s) + 2MnO_2(s) + 2NH_4^+ \rightarrow Mn_2O_3(s) + H_2O + Zn(NH_3)_2^{2+}(aq) \text{ (E}^0 = 1.5 \text{ V)}$$

Alkaline batteries are almost the same as zinc-carbon batteries, except that the electrolyte is potassium hydroxide (KOH), rather than ammonium chloride. The two half reactions in an alkaline battery are as follows:

Anode: $Zn(s) + 2OH^-(aq) \rightarrow Zn(OH)_2(s) + 2e^-$

Cathode: $2MnO_2(s) + H_2O(l) + 2e^- \rightarrow Mn_2O_3(s) + 2OH^-(aq)$

Zinc-carbon and alkaline batteries are cheap primary batteries and are therefore very useful in appliances such as remote controls, torches and radios where the power drain is not too high. The disadvantages are that these batteries can't be recycled and can leak. They also have a short shelf life. Alkaline batteries last longer than zinc-carbon batteries.

Interesting Fact

The idea behind today's common 'battery' was created by Georges Leclanche in France in the 1860's. The anode was a zinc and mercury alloyed rod, the cathode was a porous cup containing crushed MnO_2. A carbon rod was inserted into this cup. The electrolyte was a liquid solution of ammonium chloride, and the cell was therefore called a *wet cell*. This was replaced by the *dry cell* in the 1880's. In the dry cell, the zinc can which contains the electrolyte, has become the anode, and the electrolyte is a paste rather than a liquid.

23.5.5 Environmental considerations

While batteries are very convenient to use, they can cause a lot of damage to the environment. They use lots of valuable resources as well as some potentially hazardous chemicals such as lead, mercury and cadmium. Attempts are now being made to recycle the different parts of batteries so that they are not disposed of in the environment, where they could get into water supplies, rivers and other ecosystems.

?

Exercise: Electrochemistry and batteries

A dry cell, as shown in the diagram below, does not contain a liquid electrolyte. The electrolyte in a typical zinc-carbon cell is a moist paste of ammonium chloride and zinc chloride.

(NOTE TO SELF: Insert diagram)

The paste of ammonium chloride reacts according to the following half-reaction:

$$2NH_4^+(aq) + 2e^- \rightarrow 2NH_3(g) + H_2(g) \text{ (a)}$$

Manganese(IV) oxide is included in the cell to remove the hydrogen produced during half-reaction (a), according to the following reaction:

$$2MnO_22(s) + H_2(g) \rightarrow Mn_2O_3(s) + H_2O(l) \text{ (b)}$$

The combined result of these two half-reactions can be represented by the following half reaction:

$$2NH_4^+(aq) + 2MnO_2(s) + 2e^- \rightarrow Mn_2O_3(s) + 2NH_3(g) + H_2O(l) \text{ (c)}$$

1. Explain why it is important that the hydrogen produced in half-reaction (a) is removed by the manganese(IV) oxide.
   In a zinc-carbon cell, such as the one above, half-reaction (c) and the half-reaction that takes place in the Zn/Zn^{2+} half-cell, produce an emf of 1,5 V under standard conditions.
2. Write down the half-reaction occurring at the anode.
3. Write down the net ionic equation occurring in the zinc-carbon cell.
4. Calculate the reduction potential for the cathode half-reaction.
5. When in use the zinc casing of the dry cell becomes thinner, because it is oxidised. When not in use, it still corrodes. Give a reason for the latter observation.
6. Dry cells are generally discarded when 'flat'. Why is the carbon rod the most useful part of the cell, even when the cell is flat?

(DoE Exemplar Paper 2, 2007)

23.6 Summary

- The growth of South Africa's **chemical industry** was largely because of the mines, which needed explosives for their operations. One of South Africa's major chemical companies is **Sasol**. Other important chemical industries in the country are the **chloralkali** and **fertiliser** industries.
- All countries need energy resources such as oil and natural gas. Since South Africa doesn't have either of these resources, Sasol technology has developed to convert **coal** into liquid fuels.
- Sasol has three main operation focus areas: Firstly, the conversion of **coal to liquid fuel**, secondly the production and **refinement of crude oil** which has been imported, and thirdly the production of **liquid fuels from natural gas**.
- The conversion of coal to liquid fuels involves a **Sasol/Lurgi gasification** process, followed by the conversion of this synthesis gas into a range of hydrocarbons, using the **Fischer-Tropsch technology** in **SAS reactors**.
- Heavy hydrocarbons can be converted into light hydrcarbons through a process called **cracking**. Common forms of cracking are **hydrocracking** and **steam cracking**.
- With regard to crude oil, Sasol imports crude oil from Gabon and then refines this at the **Natref refinery**.
- Gas from Mozambique can be used to produce liquid fuels, through two processes: First, the gas must pass through an autothermal reactor to produce a synthesis gas. Secondly, this synthesis gas is passed through a **Sasol Slurry Phase Distillate** process to convert the gas to hydrocarbons.
- All industries have an impact on the **environment** through the consumption of natural resources such as water, and through the production of pollution gases such as carbon dioxide, hydrogen sulfides, nitrogen oxides and others.
- The **chloralkali industry** produces **chlorine** and **sodium hydroxide**. The main raw material is **brine** (NaCl).

- In industry, **electrolytic cells** are used to split the sodium chloride into its component ions to produce chlorine and sodium hydroxide. One of the challenges in this process is to keep the products of the electrolytic reaction (i.e. the chlorine and the sodium hydroxide) separate so that they don't react with each other. Specially designed electrolytic cells are needed to do this.

- There are three types of electrolytic cells that are used in this process: **mercury cell**, the **diaphragm cell** and the **membrane cell**.

- The **mercury cell** consists of two reaction vessels. The first reaction vessel contains a mercury cathode and a carbon anode. An electric current passed through the brine produces Cl^- and Na^+ ions. The Cl^- ions are oxidised to form chlorine gas at the anode. Na^+ ions combine with the mercury cathode to form a sodium-mercury amalgam. The sodium-mercury amalgam passes into the second reaction vessel containing water, where the Na^+ ions react with hydroxide ions from the water. Sodium hydroxide is the product of this reaction.

- One of the **environmental impacts** of using this type of cell, is the use of **mercury**, which is highly toxic.

- In the **diaphragm cell**, a porous diaphragm separates the anode and the cathode compartments. Chloride ions are oxidised to chlorine gas at the anode, while sodium ions produced at the cathode react with water to produce sodium hydroxide.

- The **membrane cell** is very similar to the diaphragm cell, except that the anode and cathode compartments are separated by an **ion-selective membrane** rather than by a diaphragm. Brine is only pumped into the anode compartment. Positive sodium ions pass through the membrane into the cathode compartment, which contains water. As with the other two cells, chlorine gas is produced at the anode and sodium hydroxide at the cathode.

- One use of sodium hydroxide is in the production of **soaps and detergents**, and so this is another important part of the chloralkali industry.

- To make soap, sodium hydroxide or potassium hydroxide react with a fat or an oil. In the reaction, the sodium or potassium ions replace the alcohol in the fat or oil. The product, a **sodium or potassium salt of a fatty acid**, is what soap is made of.

- The fatty acids in soap have a **hydrophilic** and a **hydrophobic** part in each molecule, and this helps to loosen dirt and clean items.

- **Detergents** are also cleaning products, but are made up of a mixture of compounds. They may also have other components added to them to give certain characteristics. Some of these additives may be abrasives, oxidants or enzymes.

- The **fertiliser industry** is another important chemical industry.

- All plants need certain **macronutrients** (e.g. carbon, hydrogen, oxygen, potassium, nitrogen and phosphorus) and **micronutrients** (e.g. iron, chlorine, copper and zinc) in order to survive. Fertilisers provide these nutrients.

- In plants, most nutrients are obtained from the atmosphere or from the soil.

- Animals also need similar nutrients, but they obtain most of these directly from plants or plant products. They may also obtain them from other animals, whcih may have fed on plants during their life.

- The fertiliser industry is very important in ensuring that plants and crops receive the correct nutrients in the correct quantities to ensure maximum growth.

- **Nitrogen fertilisers** can be produced industrially using a number of chemical processes: The **Haber process** reacts nitrogen and hydrogen to produce **ammonia**; the **Ostwald process** reacts oxygen and ammonia to produce **nitric acid**; the **nitrophosphate process** reacts nitric acid with phosphate rock to produce compound fertilisers.

- **Phosphate fertilisers** are also produced through a series of reactions. The **contact process** produces **sulfuric acid**. Sulfuric acid then reacts with phosphate rock to produce **phosphoric acid**, after which phosphoric acid reacts with ground phosphate rock to produce fertilisers such as **triple superphosphate**.
- Potassium is obtained from **potash**.
- Fertilisers can have a damaging effect on the environment when they are present in high quantities in ecosystems. They can lead to **eutrophication**. A number of preventative actions can be taken to reduce these impacts.
- Another important part of the chemical industry is the production of **batteries**.
- A battery is a device that changes chemical energy into electrical energy.
- A battery consists of one or more **voltaic cells**, each of which is made up of two half cells that are connected in series by a conductive electrolyte. Each half cell has a net electromotive force (emf) or voltage. The net voltage of the battery is the difference between the voltages of the half-cells. This potential difference between the two half cells is what causes an electric current to flow.
- A **primary battery** cannot be recharged, but a **secondary battery** can be recharged.
- The **capacity** of a battery depends on the **chemical reactions** in the cells, the **quantity of electrolyte and electrode material** in the cell, and the **discharge conditions** of the battery.
- The relationship between the current, discharge time and capacity of a battery is expressed by **Peukert's law**:

$$C_p = I^k t$$

  In the equation, 'C_p' represents the battery's capacity (Ah), I is the discharge current (A), k is the Peukert constant and t is the time of discharge (hours).
- Two common types of batteries are **lead-acid batteries** and the **zinc-carbon dry cell**.
- In a **lead-acid battery**, each cell consists of electrodes of lead (Pb) and lead (IV) oxide (PbO_2) in an electrolyte of sulfuric acid (H_2SO_4). When the battery discharges, both electrodes turn into lead (II) sulphate ($PbSO_4$) and the electrolyte loses sulfuric acid to become mostly water.
- A **zinc-carbon cell** is made up of an outer zinc container, which acts as the **anode**. The cathode is the central carbon rod, surrounded by a mixture of carbon and manganese (IV) oxide (MnO_2). The electrolyte is a paste of ammonium chloride (NH_4Cl). A fibrous fabric separates the two electrodes, and a brass pin in the centre of the cell conducts electricity to the outside circuit.
- Despite their many advantages, batteries are made of potentially toxic materials and can be damaging to the **environment**.

Exercise: Summary Exercise

1. Give one word or term for each of the following descriptions:
   (a) A solid organic compound that can be used to produce liquid fuels.
   (b) The process used to convert heavy hydrocarbons into light hydrocarbons.
   (c) The process of separating nitrogen from liquid air.
   (d) The main raw material in the chloralkali industry.
   (e) A compound given to a plant to promote growth.

(f) An electrolyte used in lead-acid batteries.

2. Indicate whether each of the following statements is true or false. If the statement is false, rewrite the statement correctly.
   (a) The longer the hydrocarbon chain in an organic compound, the more likely it is to be a solid at room temperature.
   (b) The main elements used in fertilisers are nitrogen, phosphorus and potassium.
   (c) A soap molecule is composed of an alcohol molecule and three fatty acids.
   (d) During the industrial preparation of chlorine and sodium hydroxide, chemical energy is converted to electrical energy.
3. For each of the following questions, choose the one correct answer from the list provided.
   (a) The sequence of processes that best describes the conversion of coal to liquid fuel is:
      i. coal → gas purification → SAS reactor → liquid hydrocarbon
      ii. coal → autothermal reactor → Sasol slurry phase F-T reactor → liquid hydrocarbon
      iii. coal → coal purification → synthesis gas → oil
      iv. coal → coal gasification → gas purification → SAS reactor → liquid hydrocarbons
   (b) The half-reaction that takes place at the cathode of a mercury cell in the chloralkali industry is:
      i. $2Cl^- \rightarrow Cl_2 + 2e^-$
      ii. $2Na^+ + 2e^- \rightarrow 2Na$
      iii. $2H^+ + 2e^- \rightarrow H_2$
      iv. $NaCl + H_2O \rightarrow NaOH + HCl$
   (c) In a zinc-carbon dry cell...
      i. the electrolyte is manganese (IV) oxide
      ii. zinc is oxidised to produce electrons
      iii. zinc is reduced to produce electrons
      iv. manganese (IV) dioxide acts as a reducing agent
4. **Chloralkali manufacturing process**
   The chloralkali (also called 'chlorine-caustic') industry is one of the largest electrochemical technologies in the world. Chlorine is produced using three types of electrolytic cells. The simplified diagram below shows a membrane cell.

   (a) Give two reasons why the membrane cell is the preferred cell for the preparation of chlorine.
   (b) Why do you think it is advisable to use inert electrodes in this process?
   (c) Write down the equation for the half-reaction taking place at electrode M.
   (d) Which gas is chlorine gas? Write down only Gas A or Gas B.

(e) Briefly explain how sodium hydroxide forms in this cell.

(DoE Exemplar Paper 2,2007)

5. The production of nitric acid is very important in the manufacture of fertilisers. Look at the diagram below, which shows part of the fertiliser production process, and then answer the questions that follow.

(a) Name the process at (1).
(b) Name the gas at (2).
(c) Name the process at (3) that produces gas (2).
(d) Name the product at (4).
(e) Name two fertilisers that can be produced from nitric acid.

6. A lead-acid battery has a number of different components. Match the description in Column A with the correct word or phrase in Column B. All the descriptions in Column A relate to lead-acid batteries.

Column A	**Column B**
The electrode metal	Lead sulphate
Electrolyte	Mercury
A product of the overall cell reaction	Electrolytic
An oxidising agent in the cathode half-reaction	Lead
Type of cells in a lead-acid battery	Sulfuric acid
	Ammonium chloride
	Lead oxide
	Galvanic

Appendix A

GNU Free Documentation License

Version 1.2, November 2002
Copyright © 2000,2001,2002 Free Software Foundation, Inc.
59 Temple Place, Suite 330, Boston, MA 02111-1307 USA
Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

PREAMBLE

The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or non-commercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "Document", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "you". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "Modified Version" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "Secondary Section" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "Invariant Sections" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The "Cover Texts" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A "Transparent" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called "Opaque".

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The "Title Page" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

A section "Entitled XYZ" means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as "Acknowledgements", "Dedications", "Endorsements", or "History".) To "Preserve the Title" of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or non-commercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section A.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts,

you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections A and A above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

1. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.
2. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.
3. State on the Title page the name of the publisher of the Modified Version, as the publisher.
4. Preserve all the copyright notices of the Document.
5. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.
6. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.
7. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.
8. Include an unaltered copy of this License.
9. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.

10. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.

11. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.

12. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.

13. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.

14. Do not re-title any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.

15. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties–for example, statements of peer review or that the text has been approved by an organisation as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section A above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements".

COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section A is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section A. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section A) to Preserve its Title (section A) will typically require changing the actual title.

TERMINATION

You may not copy, modify, sub-license, or distribute the Document except as expressly provided for under this License. Any other attempt to copy, modify, sub-license or distribute the Document is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See http://www.gnu.org/copyleft/.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation.

ADDENDUM: How to use this License for your documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

> Copyright © YEAR YOUR NAME. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License".

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the "with...Texts." line with this:

with the Invariant Sections being LIST THEIR TITLES, with the Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.