PHYSICS FOR BEGINNERS

A Novice's Guide to the Mysteries of the Universe

by

Matthew Raspanti

Self-published books by Matthew Raspanti, available at amazon.com:

The Virtual Universe – Philosophy, Physics and the Nature of Things (1998)

Virtualism, Mind and Reality – An Approach to Untangle the Consciousness Problem (2008)

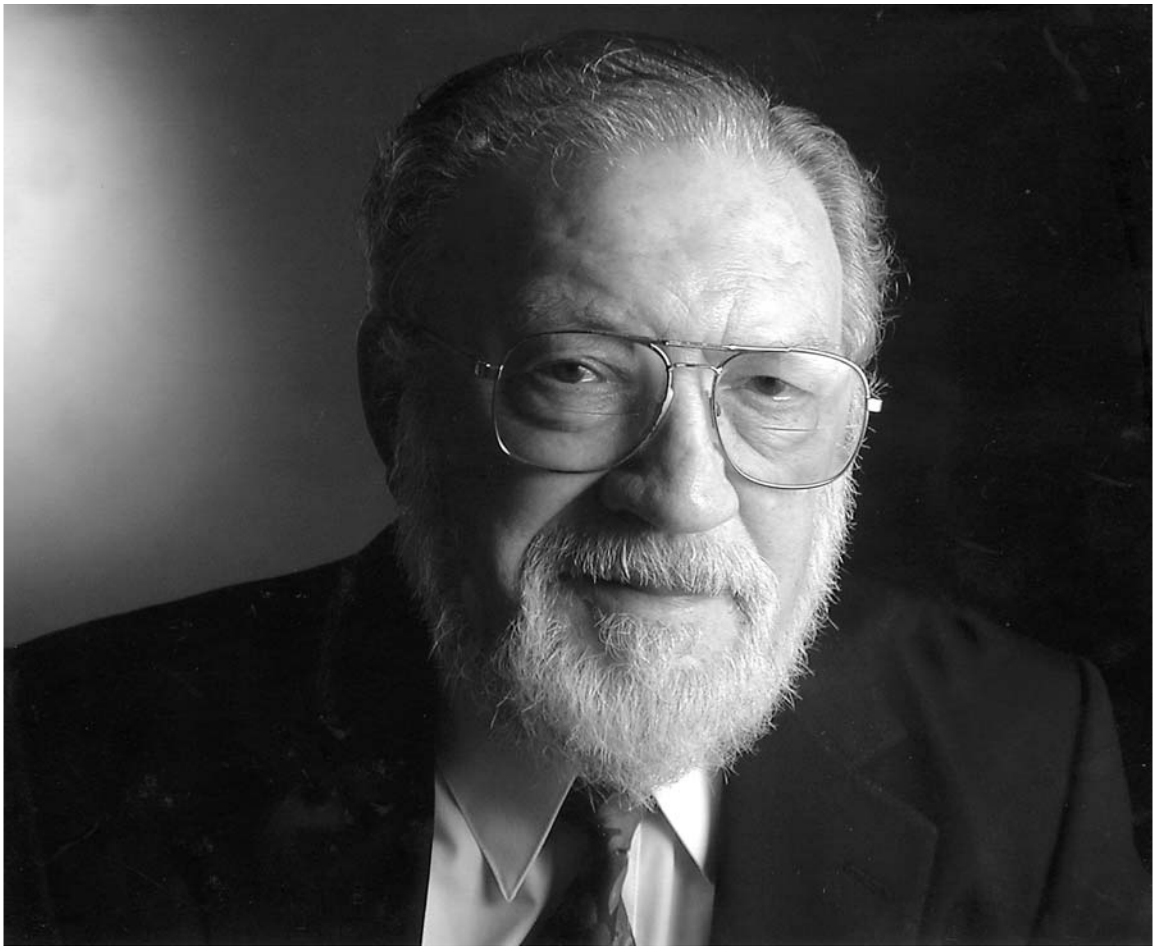

Photo by danny sanchez photographer, Red Bank, NJ

11/20/08

Matthew Raspanti

CONTENTS

PREFACE

I was born in 1924 in New York City. When I was seven, however, my family moved to Sicily. I lived there until I graduated from the University of Palermo with a doctorate in industrial engineering summa cum laude. After returning to the States in 1947, I earned a master's degree in electrical engineering from the Polytechnic Institute of Brooklyn. Starting in 1954, I was a member of the technical staff at world-renowned Bell Labs for 35 years; for the last 24, as a department head. My fields of interest were the hardware and software of computers and computer-controlled telephone switching systems, fields in which I did also part-time college-level teaching for a number of years. I hold three patents, and am an associate member of the research fraternity Sigma Xi. I retired in 1990.

For many years, I have been a great admirer of physics, its quest, methods and achievements. After retiring, I decided I would revisit physics both for my own sake and to write about it in a book for lay readers. I tried to have the book published but without success. I then moved on to other projects that had been sparked in the meantime by my writing. The book lay dormant for years in my computer hard-drive. I distributed copies to a few people, and their reactions confirmed my own expectation that the book can be very helpful to a beginner.

To make the book available to as many interested people as possible, I have decided to offer it free in digital form on the Internet. I have e-mailed copies to several people, and plan to make it available at a website of mine, where it will be downloadable at no cost. The book can be freely printed for personal use, or forwarded to others.

I will greatly appreciate any feedback, most particularly if something is found that would be unacceptably wrong, even in a book for beginners, where a few “poetical” licenses are unavoidable, or even desirable for the sake of clarity.

Matthew Raspanti
mraspanti.ph@verizon.net

November 20, 2008

Chapter 1

INTRODUCTION

Most people know, at least in some vague way, that the sophisticated technology that drives our society has been driven in turn by fundamental discoveries of physics. But, just what is physics? It derives its present name from the Greek word for nature; it was previously called natural philosophy. Physics can be defined as the science that deals with matter, energy, motion and force. It studies the fundamental building blocks of the universe and how they interact. It seeks answers to such fundamental questions as: What kind of world do we live in? How does it work? What are the fundamental laws of nature? Thus, physics is the basic science from which all others have derived.

Transistors, microchips, lasers, computers, telecommunications, nuclear power and space travel are among the many applications of physics that are so pervasive in our times. In our daily newspaper or weekly magazine, we often find articles that attempt to explain to a lay public a variety of topics related to physics. These might be sophisticated experiments on fundamental particles of matter; space probes and their missions; discoveries of astronomy in very remote regions of space; exotic new theories on the nature of matter, or the universe as a whole.

The relevance of physics is all around us. Although not as palpable as in the days of the Cold War with the Soviet Union, the terrifying threat of nuclear holocaust still hangs over all mankind. With so many programs competing for federal funds, government support of very expensive scientific ventures has become an issue of public interest. Except for fundamentalist groups, few, if any, religious leaders dare challenge the experimental findings of physics. No metaphysical speculation about the nature of reality[1], whether by lay people or professional philosophers, can ignore these findings. We clearly live in times that require at least some modest level of literacy in physics, one of the most profound achievements of the human mind. Unfortunately, physics is the least known and the most intimidating of all sciences. This is true even for many who are literate at some level about other human endeavors.

Among the factors that make physics appear so alien to so many people are the difficulty of many of its concepts, its pervasive use of advanced mathematics and cryptic symbolism, and the sophistication of its instruments, whose complexity goes far beyond the telescope first used by Galileo in 1609.

Although strongly intimidated by physics, much of the lay public has

---

[1] In this book, the terms world, universe and reality will be used interchangeably. The term "reality" derives from the Latin word "res" meaning thing. Thus, reality refers to the totality of all things.

been, and still is, intrigued by the fundamental nature of its inquiry. This is shown by the success of dozens of books that have been written since Stephen Hawking' s "A Brief History of Time" (1988) became a best seller.

In most of the popular books on the market, however, the bulk of the material is at a level of presentation and detail that goes beyond the background and interest of much of the general public. (A notable exception is Roger S. Jones' very readable "Physics for the Rest of Us", Contemporary Books, 1992). Many of these books focus on specific areas of scientific endeavor; some are offered as part of a series that covers a broader area of physics.

This book is devoted to a basic, non-mathematical presentation of physics to *motivated beginners*, that is, intelligent people who have no prior scientific or mathematical background, but are interested in learning something about this fundamental science. While many may not wish to go beyond this book, others could profitably use it as the first stepping stone to more advanced popular books.

Physicists undergo a long and demanding training in order to be able to do their work. It is far from hopeless, however, for a motivated beginner to acquire some general, conceptual understanding of many of physics' basic ideas and their philosophical significance.

In a concise, straightforward and reader-friendly style, the book presents an overview of physics in semi-historical sequence, enlivened here and there by biographical sketches of some of the major players. The semi-historical style makes possible a gradual presentation of new concepts, each supported by the necessary background. This style has also the advantage of giving some sense of how some of the greatest scientific discoveries gradually unfolded over the centuries. The book can then be seen as a brief history of the human quest for answers to the mysteries of the universe.

There is no way that a book on physics can be written to read like a novel. The motivated reader, however, may come to see the story of physics as an intriguing detective novel, in which a side detail of one day becomes a crucial clue to a later discovery.

A book that attempts to popularize a subject as complex as physics faces the obvious necessity of omitting all of the math and most of the material. Much more difficult is deciding where to dwell more deeply (without losing the reader) and where to go more lightly (without trivializing the material). Inevitably, from one section to another, what is too much for one reader will be not enough for another.

The book begins with ancient astronomy and the laws of motion, and then leads the reader through the essentials of energy, atoms, molecules, particles in motion, waves, light, and electromagnetism (Chapters 2-9). These all serve as preliminaries to the two fundamental theories on which contemporary physics rests: relativity and quantum mechanics (Chapters 10-17). Chapter 18 gives a summary of the fundamental particles of matter and the forces by which they interact. Chapter 19, the last, gives a cosmic perspective based on the currently prevailing theories on the origin, evolution, structure and future of the universe as a whole.

Chapter 2

HEAVENLY MOTIONS

Modern physics - as the systematic study of nature based on observation, experimentation, reason and mathematical analysis - had its beginnings in the 1600's with the work of Galileo Galilei and Isaac Newton. Their accomplishments, however, owed much to earlier discoveries. In fact, mankind's efforts to understand nature can be traced back thousands of years.

EARLY ASTRONOMY

Astronomy was the first science to emerge; for thousands of years, it was the "Queen of sciences." Since ancient times, mankind has been awed and intrigued by the grand canvass of the sky with its myriad stars, planets and comets, and by the motions of these heavenly bodies.

The ancients believed that the heavens were inhabited by gods, and that heavenly phenomena could influence earthly events. Thus, religion, astrology and astronomy were intimately linked: a combination of the three can be found in the early history of places as diverse as ancient Mesopotamia (present-day Iraq), Egypt, China, India and Central America.

The heavens, which presented an irresistible puzzle to the curiosity of early humankind, eventually became the source of very useful knowledge. Early on, it was observed that the stars in the night sky did not seem to move with respect to one another, but appeared fixed in a pattern that rotated daily about the Earth. Later, nomads discovered that they could be guided in their travels by their familiarity with certain clusters of stars, or constellations. Later still, when nomads settled down and became farmers, knowledge of the constellations helped them keep track of the seasons.

Very early, it was noticed that, beside the Sun and the Moon, a few heavenly bodies moved against the background of the fixed stars in the course of a year. Only seven such wandering bodies were known to the ancients: the Sun, the Moon, Mercury, Venus, Mars, Jupiter and Saturn. Later, they were all called "planets", from the Greek word for "wanderers."

In the familiar seven-day week, which was first introduced by the ancient Babylonians in Mesopotamia, the names of the days can be traced in various languages to the seven planet-gods. In English, Saturday, Sunday and Monday are associated with Saturn, the Sun and the Moon, respectively. In Romance languages, such as French or Italian, the names of the days from Monday through Friday are associated with the Moon, Mars, Mercury, Jupiter and Venus, respectively.

Foremost among the "wanderers" was the Sun. Its rising and setting defined the cycle of day and night. Its yearly motion through various constellations defined the time to seed and the time to harvest. Next in importance was the Moon, whose shape could be seen to change through four regularly repeated phases.

By observing the motions of the Sun and the Moon, the ancients were able to develop various calendars based on solar or lunar cycles. These calendars made it possible to set the times for religious ceremonies, farming and other events. Together with the Sun and the Moon, the other five planets played an important role in astrology.

Western astronomy had its origins in ancient Egypt and Mesopotamia. In Egypt, astronomy was concerned primarily with predicting the time of the annual flooding of the river Nile, which played a crucial role in the life of that country by fertilizing its land. Egyptian astronomy's main lasting contribution was a calendar of 365 days, divided into 12 months of 30 days each, with five feast days at the end of the year.

In Mesopotamia, powerful and capricious gods were believed to inhabit the skies. Since heavenly phenomena were deemed to foretell earthly disasters[2], they were carefully observed and recorded. These practices led to a highly developed mathematics and the most sophisticated astronomy in the ancient world until the Greeks. It was the Greeks who first attempted to explain natural phenomena on the basis of reason, rather than the arbitrary will of gods. They were also the first to apply geometry to astronomy.

GREEK CIVILIZATION

The classical period of Greek civilization started in the 5th century BC. With their prodigious creativity in art, architecture, literature, philosophy and science, the Greeks laid the foundations of Western culture.

The area in which this civilization flourished extended beyond mainland Greece and the islands of the Aegean Sea. It included also the many colonies the Greeks had established all along the western coast of present-day Turkey, and in parts of southern Italy and Sicily. These lands were collectively referred to as "Hellas", and the Greeks as "Hellenes".

From Macedonia, which bordered on northern Greece, in the 4th century BC, the Greek-educated Alexander the Great (356-323 BC) conquered a vast empire, which included Greece and Egypt, and extended from Turkey and the Middle East to western India. The three centuries after Alexander's death - known as the Hellenistic period - were among the most creative in all of Greek history. In Egypt, the city of Alexandria became a

---

[2] The word *disaster* derives from *dis-* (opposite) and *aster* (star).

leading center of Greek scholarship: its famous library held more than 500,000 scrolls, before it was destroyed by fire.

The systematic application of reason to the explanation of natural phenomena represents the most enduring legacy of Greek culture to modern science. It led to major accomplishments in philosophy, geometry and astronomy, and reached its peak with the philosopher Aristotle (384-322 BC).

Early Greek Philosophy

In its very beginnings, Greek philosophy was concerned with nature. This early "natural philosophy" was the forerunner of modern physics (which was still called natural philosophy well into 1800's). Starting in the 6th century BC, the earliest Greek philosophers theorized that, beneath the great variety of nature, all things were made of only a few fundamental substances, or "elements". By the 4th century BC, most philosophers supported a theory of only four elements: earth (soil), water, air and fire.

A totally different conception of reality was held by the "Atomists" (5th century BC). They believed that the physical world consisted of countless, unchangeable, indivisible particles, which differed in size and shape and moved in a "void", or vacuum (empty space). These fundamental particles were called "atoms" from the Greek word for indivisible. By combining in various ways, they formed all the matter in the universe. All natural phenomena resulted from the variety of motions and configurations of atoms in empty space.

Greek Geometry

Geometry developed in Egypt and Mesopotamia as an empirical art based on practical experience. Later, Greek geometers gradually transformed it into a logical system.

The most famous book on Greek geometry is "The Elements" by Euclid, who taught in Alexandria about 300 BC. It has been said that, next to the Bible, Euclid's Elements may have been the most translated and studied book in the Western world. Euclid presented his subject using the so-called axiomatic-deductive method, which has since served as the model for the development of many other mathematical subjects.

In the Elements, after some basic definitions, Euclid introduces ten statements offered as "axioms" or "postulates", i.e., to be accepted as true without proof, because they are deemed self-evident. Step by step, he proceeds then to prove a number of "theorems", each built on postulates and previous theorems. Thus, starting from statements that are accepted as immediately evident, the reader is led, by a long series of logical steps, to accept the truth of much more complex statements, which would not have been accepted at the start.

In the Elements, Euclid gave a compilation of all the geometry that had been developed during the preceding two centuries. Most of the discoveries presented were by earlier geometers. Euclid's major contribution was mainly the excellent organization of the subject matter.

In the third century BC, Greek geometry entered its golden age, which was dominated by the discoveries of two men: Archimedes and Apollonius.

Archimedes of Syracuse (Sicily, 287?-212 BC) discovered how to compute the area and volume of the sphere and many other complex shapes. He also showed how the number "pi", the ratio of the circumference of a circle to its diameter, could be computed to any desired accuracy. His approximation of 3.14 for "pi" was used well into the Middle Ages. Archimedes' writings deeply influenced later mathematicians and scientists, most notably Galileo and Newton.

Apollonius of Perga (southwestern coast of Turkey, 262-190 BC) studied and taught in Alexandria. He came to be known by his contemporaries as the Great Geometer. His main surviving work, "Conics", is among the greatest scientific works from the ancient world. Conics are a family of geometric curves, so called because they can all be generated by cutting across a cone in various ways. Among them are the familiar circle, the ellipse and the parabola, about which more will be said later. As we will see, conics have played an important role in physics. They provide one of many instances of knowledge initially pursued for its own sake, and later found very useful for practical as well as theoretical purposes.

Greek Astronomy

Greek astronomers were the first to use geometry to develop their field into a science. Like their predecessors, they believed in a geocentric (Earth-centered) universe: the Sun, the Moon and the five known planets were all believed to revolve around the Earth, which stood still at the center of a rotating sphere to which all the stars were attached in a fixed pattern.

Some heliocentric (Sun-centered) proposals were made, but were rejected because the notion of a moving Earth seemed totally contrary to common-sense intuition. In the 4th century BC, Heracleides was the first to maintain that the Earth rotates about its axis, and that Mercury and Venus (the two planets between the Sun and the Earth) revolve around the Sun. Aristarchus is believed to have been the first to propose, in the 3rd century BC, a completely heliocentric theory similar to the one that Copernicus was to propose in the 16th century.

To describe the motions of the seven "planets", Greek astronomers used only circular motions or combinations of circular motions. In this, they were influenced by the teachings of the great philosopher Plato (428-348

BC). He believed that heavenly bodies were divine and, therefore, perfect. As such, they had to be endowed with perfect motion which, by Plato's definition, was circular uniform motion (i.e., motion along a circle at constant speed).

Greek astronomers viewed their geometric schemes simply as tools for predicting planetary positions. They did not know how the "planets" actually moved, or why.

An earlier scheme was based on a complex system of interconnected "homocentric" (concentric) spheres, all nested inside one another. The Earth stood still at the common center. The "planets" rode on the equators of seven of the spheres. It is this scheme that was incorporated by Aristotle in his "System of the World", a grand account of the whole Universe.

Aristotle's "System of the World"

No philosopher has influenced Western thought more than Aristotle. His intellectual interests covered most of the sciences and many of the arts. His greatest achievement was his system of formal logic, which for centuries represented the totality of logic. Aristotle's philosophical and scientific system became the foundation of both Christian and Islamic thinking in the Middle Ages. Until the 17th century, Western culture was strongly Aristotelian, and even today many Aristotelian concepts remain embedded in Western thinking.

Aristotle was born in 384 BC in Northern Greece; his father was court physician to the king of Macedonia. He studied in Athens under Plato for 20 years. After Plato's death, he traveled for 12 years. During this period, he spent three years at the court of Macedonia as the tutor of the king's son, who would become known as Alexander the Great. After returning to Athens, he opened a center for studies in all fields.

Aristotle viewed the universe or "cosmos" as an ordered structure. Indeed, the word cosmos comes from the Greek word for order. This order was believed to be that of an organism: all parts of the universe had purposes in the overall scheme of things, and objects moved naturally toward the ends they were intended to follow.

Aristotle adopted the system of homocentric spheres as the actual physical machinery of the heavens. His cosmos was like an onion consisting of "crystalline" (transparent) spheres all nested around the Earth. The ultimate cause of all motion was a Prime Mover (God), who stood outside the cosmos.

Aristotle's universe is divided into a terrestrial, or earthly, realm and a celestial, or heavenly, realm. All terrestrial objects are made of one or more of four elements: earth, water, air and fire. Each element has its assigned natural place: earth at the center is surrounded by concentric spheres of water, air and fire. The fixed stars, the Sun, the Moon and the planets, which move in the celestial region, are all composed of a fifth essence, or quintessence, called "ether".

Different laws govern the celestial and terrestrial realms. In the celestial realm, uniform circular motion is the natural form of motion. In the terrestrial realm, instead, natural motion is either up or down. Light bodies (like smoke), by their nature, tend to move up. On the other hand, heavy bodies, by their nature, seek the center of the universe, and tend to move down. When a body (object) falls freely through air, the heavier the body, the faster it falls.

An external cause is needed to put a body in motion. As long as the body moves, a force must be in constant, direct contact with it, causing it to move. This theory, however, does not satisfactorily explain why a stone thrown from a sling continues to move up before it starts falling down.

Aristotle's "system of the world" was a magnificent attempt to unify all the branches of human knowledge within a single conceptual framework. After the fall of the Roman Empire, Aristotle's works (like those of many ancient authors) were lost in the West; they were preserved, however, by Arabic and Jewish scholars. Muslim scholars kept alive the Aristotelian heritage, and in the 12th and 13th centuries passed it back to Europe, where it became the philosophical basis of Christian theology.

Epicyclic Motions

After Aristotle, Greek astronomers abandoned the scheme of homocentric spheres and adopted a new geometric scheme, which reflected two phenomena observed in the motions of planets. One was the fact that planets could be observed to change in brightness, which suggested that they were not always at the same distance from the Earth.

The other phenomenon was the puzzling to-and-fro motion of a planet, called "retrograde" motion. When viewed against the background of the fixed stars at the same hour on many consecutive nights, the motion of a planet appears to be generally eastward, with occasional stops followed by temporary reversals to a westward direction, until the planet appears to stop again and then resume its eastward motion.

Both retrograde motion and changes in brightness could be accounted for by the scheme of "epicyclic" motions, as illustrated in **Figure 2.1**. (For this figure, as well as most of the others, the text discussing the figure appears right under it, in Italics. In general, skipping a figure and its description will result in loss of continuity.)

A major contributor to the theory of epicycles was Hipparchus of Nicaea (northwestern Turkey, 190-120 BC), who was, most likely, the greatest astronomer of antiquity. Among his achievements was a catalog of stars, the first ever to be compiled. Started in 134 BC and completed 5 years later, this catalog listed about 850 stars classified by brightness.

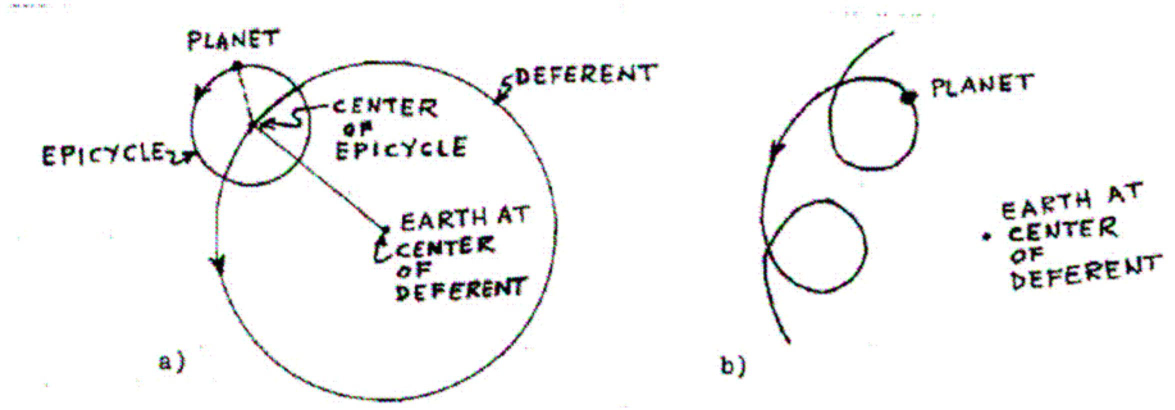

Figure 2.1 - Epicyclic motion

As shown on the left, a planet (represented as a point) is imagined to move along a circle, the "epicycle", whose center in turn moves along a larger circle, the "deferent". The latter's center is at the center of the Earth. The resulting motion of the planet is shown on the right. There is a back-and-forth retrograde motion combined with a motion toward, or away from, the Earth.

The Ptolemaic System

Although most works of Greek astronomers were lost, their contents are known primarily through a major book on astronomy by Claudius Ptolemy of Alexandria (100-170 AD). This book, originally called "The Mathematical Composition", eventually became better known as the "Almagest (The Greatest)", from the title of its Arabic translation. It provided a compendium of Greek astronomy, comparable in thoroughness to Euclid's compendium of geometry. It shaped astronomy for 1400 years until Copernicus.

Expanding on Hipparchus' work, Ptolemy formulated what is popularly known as the Ptolemaic system. In the first part of the Almagest, Ptolemy describes the system, and presents various arguments to prove that the Earth must be standing still at the center of the universe. Since, according to Aristotelian physics, all bodies tend to fall toward the center of the universe, the Earth must be fixed there, otherwise objects would not be observed to fall toward the Earth. Also, if the Earth rotated once every 24 hours, as claimed by some, a body thrown vertically up would not be observed to fall back to the same place. On the strength of such arguments, geocentrism was eventually elevated to the status of almost religious dogma.

Like his predecessors, Ptolemy placed the seven heavenly bodies in the following order, starting from the Earth: Moon, Mercury, Venus, Sun, Mars, Jupiter and Saturn. The detailed geometric models he adopted for each of the seven "planets" fitted the available data with sufficient accuracy to give good validity to the system.

FROM PTOLEMY TO COPERNICUS

Without major breakthroughs, Western science continued to operate within the framework of Aristotle's physics and Ptolemy's astronomy for some 1400 years until the scientific revolution that started in the 16th century with Copernicus.

The Roman Empire

After the Romans conquered Greece, they were greatly impressed by the intellectual achievements of the Greeks, but doubtful about their practical value. The Greeks' pursuit of knowledge for its own sake was alien to the practical Roman mind (whose greatest legacy, instead, was jurisprudence, the philosophy of law). As a result, scientific innovation came to a halt under the Romans.

By Ptolemy's time in the 2nd century AD, the Roman Empire had reached its greatest extent to include all the lands bordering on the Mediterranean. Eventually, the empire became so unwieldy that, in 395 AD, it was split in two. The Western Empire lasted less than another century: in 476 AD, Rome fell as Western Europe was overrun by barbaric tribes. The Eastern Empire became known as the Byzantine Empire from its capital, the ancient city of Byzantium, later renamed Constantinople and now called Istanbul. The Eastern Empire lasted another thousand years until it fell in 1453 under the attack of the Ottoman Turks, who originated from central Asia.

The Middle Ages

After the fall of Rome, ancient learning barely survived in the West. Its basics continued to be taught in monasteries; its surviving works were faithfully copied by monks. In the Byzantine Empire, the ancient traditions continued, but little original work was done in science.

The torch of scholarship passed to the Arabs, who made major contributions to mathematics (including the invention of algebra) and built great astronomical observatories. In the 7th century, inspired by Islam, the new religion founded by Mohammed, the Arabs launched the conquest of a great empire, which eventually extended toward India and, to the West, included Northern Africa and Spain.

In Spain, when the Moslem conquerors were gradually pushed south by the Christian armies, among the treasures they left behind were Arabic translations of Greek works of science and philosophy. In 1085, the city of Toledo, with one of the finest libraries in the Islamic world, fell to the Christians. Soon, Christian monks started translating ancient works from Arabic into Latin. By the end of the 12th century, much of the ancient heritage was again available in the West, bringing about a revival of Greek science.

After the fall of the Western Empire, the Church became the intellectual, as well as the spiritual, center of Western Europe. During most of the Middle Ages, classical studies and virtually all intellectual endeavors were pursued by members of various religious orders.

Medieval thinking culminated in the philosophical system of "Scholasticism", whose leading exponent was Thomas Aquinas (Italian Dominican monk and theologian, 1224-1274). Borrowing freely from Aristotle's philosophy, the Scholastics made a systematic attempt to establish theology as a science, in which reason played an important role, not as the opponent of faith, but as its supplement.

The Renaissance

The Middle Ages were followed by the Renaissance (rebirth), a period that saw a great revival of classical learning and values. This led to cultural achievements of an extraordinary nature. Some historians define the Renaissance as the period extending roughly from the mid-1300's to the early 1600's.

The Renaissance started with the intellectual movement called Humanism, which originated in Italy. Humanism was started by lay men of letters, rather than the clerical scholars who had dominated intellectual life in the Middle Ages.

Humanism was greatly propelled by the fall of Constantinople to the Turks in 1453, as many scholars fled to Italy, bringing with them important books and a tradition of Greek scholarship.

Humanism derived its name from the Latin word "humanitas", a term used to convey the Greek ideal of education as the full development of the best in human nature. As scholars and teachers, the Humanists promoted first-hand reading of the classics and the study of such disciplines as grammar, rhetoric, poetry, moral philosophy and history (we still call them the humanities).

Humanism exalted the dignity of man as the center of the universe, with unlimited capacities for development. Instead of the medieval ideal of a life of penance as the noblest form of existence, the Humanists glorified creativity and mastery over nature.

From Italy, Humanism and the Renaissance it created spread north to all parts of Europe, greatly aided by the invention of the printing press, which increased enormously the availability of classical texts.

While generally faithful to Christian beliefs, Humanism inspired scientific inquiry free of the constraints of rigid religious orthodoxy.

The Reformation

Over the centuries, the Church's increasing power and wealth, combined with its deep involvement in the political intrigues of Western Europe, led to its spiritual deterioration. Corruption and abuses within the Church led to the Reformation movement, which at first demanded reform and ultimately chose separation, leading to the establishment of Protestantism.

The Reformation started in Germany in 1517 when the Augustinian monk Martin Luther publicly protested against the corruption of the Church, and later denied the authority of the Pope. The political and religious repercussions were enormous. The Church's response was the Counter-Reformation movement for the elimination of corruption, the reassertion of papal authority, and the suppression of heresy.

PRE-GALILEO ASTRONOMY

Copernicus (1473-1543)

Nicolaus Copernicus was the Latin name of Mikolaj Kopernik, born in East Prussia, in a region that is now part of Poland. After his father died, he was raised by an uncle, a priest, who later became an influential bishop. In 1496, he was sent to study in Italy. During the ten years he spent there, he studied both law and medicine. He also became well versed in Greek, mathematics and, particularly, astronomy.

After returning to his country, he served for six years as physician to his uncle the bishop until the latter's death. He then assumed his duties as a canon of the wealthy cathedral of Frauenburg, a well-paid post his uncle had secured for him years before. While involved in his daily occupation, he continued to pursue his interest in astronomy.

In the course of his studies, Copernicus had become increasingly dissatisfied with the Ptolemaic system. Over the centuries after Ptolemy, as astronomical observations became more accurate, in order to achieve a better match between the geometric models used and the observed positions of the planets, adjustments were made by adding epicycles riding on other epicycles, until the system became very cumbersome. Right after his return from Italy, Copernicus started developing his own system.

He proposed that it is the Sun that is fixed at the center of the universe, and that the Earth and the other planets revolve around the Sun. The Earth also rotates about its axis, thus accounting for the apparent motion of the fixed stars. Starting from the Sun, he correctly adopted the following order for the planets: Mercury, Venus, Earth, Mars, Jupiter and Saturn.

In his system, Copernicus placed all the centers of planetary orbits at the center of the Sun. More faithful to Plato's doctrine than Ptolemy had been, Copernicus strongly believed that the planets had to follow true uniform circular motion. Since the planets, however, do not move in this manner, he too was forced to resort to epicycles and other corrections. Ironically, Copernicus' system ended up requiring 48 separate circular motions, whereas the Ptolemaic system had required only 40.

In 1533, Copernicus gave lectures on the basic principles of his system before the Pope, who gave his approval. Three years later, he received a formal request to publish his thoughts. Probably in 1530, he had completed the manuscript of his book, "On the Revolutions of the Celestial Spheres", which was written in Latin, at the time the international language of scholars. Only in 1540, however, at the urging of friends, he agreed to submit the book for printing. Supposedly, Copernicus first saw an advance copy of his book on his deathbed, in 1543.

The main advantage of Copernicus' system was its ability to explain the puzzling to-and-fro "retrograde" motion of the planets. Let us consider, for instance, the Earth and Mars, as they both revolve around the Sun. As the Earth passes closest to Mars and then overtakes it, Mars - as seen from the Earth - appears to slow down, stop and then go backwards for a while.

In the view of most contemporary scholars, however, this advantage did not offset the formidable objections it raised, such as:

- According to Aristotelian-Scholastic philosophy, the Earth - the home of Man created in God's image - was located at the center of the universe. In Copernicus' system, the Earth was just one more planet revolving around the Sun.
- The new system appeared to contradict the Bible. As Luther remarked, "the sacred scriptures tell us that Joshua commanded the Sun to stand still, not the Earth".
- According to Aristotelian physics, bodies fell to the ground because they were seeking their natural place at the center of the universe. A new explanation was needed if the Earth was not at the center of the universe.
- The notion of a moving Earth was very difficult to reconcile with every-day experience. It was felt that, if the Earth were moving through space at great speed (actually, about 19 miles per second), everything on it would surely be

swept off its surface.

Because of these objections and the fact that Copernicus' system offered no advantage of simplicity with respect to the Ptolemaic system, the new system was largely ignored for decades.

Tycho Brahe (1546-1601)

A compromise between Ptolemy and Copernicus was proposed by a Danish astronomer, Tycho Brahe. He was born in a noble family three years after Copernicus' death. He studied law to please his family, but his true passion was astronomy.

He became a recognized astronomer at age 27, after reporting his discovery of a new star. In 1576, King Frederick II of Denmark gave him generous financial support to establish a large observatory. Tycho devoted great care to the design of his instruments, and lavished money to secure the finest materials and the best craftsmen. (These were all naked-eye instruments; the first use of the telescope was made by Galileo in 1609, eight years after Tycho's death.)

The mission Tycho undertook for himself and his assistants was to measure the position of every visible star with the highest possible precision, and to achieve the same accuracy in very frequent measurements of planetary positions over long periods of time. All previous astronomers had been satisfied to make only occasional observations.

After the death of Frederick II in 1588, royal support greatly diminished. Nine years later, Tycho left Denmark and, in 1599, he settled in Prague as official astronomer to the Holy Roman Emperor.

With the extensive data accumulated during the 21 years spent at his observatory in Denmark, Tycho felt ready to develop his own system, the Tychonic system, in which the five known planets revolved around the Sun, while the Sun and the Moon revolved around the Earth. This was an appealing compromise since it kept the Earth still at the center of the universe.

To develop his system, Tycho needed an assistant highly skilled in mathematics. In 1600, he hired a young German astronomer, Johannes Kepler.

Johannes Kepler (1571-1630)

Kepler came from a poor family, but a superior intelligence earned him scholarships that enabled him to study philosophy, mathematics and astronomy. After receiving his college degree in 1591, he began to study theology with the intention of becoming a Lutheran minister. In 1594, however, he was offered a post to teach mathematics. Reluctantly, he accepted it at the urging of his professors. For a number of years, he supplemented his income by publishing an astrological calendar. Throughout his life, his skill at astrological prediction was in great demand.

In 1597, Kepler published a book, which displayed his skills as a mathematician and his knowledge of astronomy. They impressed Tycho Brahe, who offered him a job.

When Tycho died a year later, Kepler at 30 was appointed his successor as imperial mathematician. He also inherited Tycho's data - but only after a considerable struggle with the heirs. This was the best collection of astronomical data that had ever been assembled.

As a member of Tycho's team, Kepler's first assignment had been to calculate an orbit that would describe the position of Mars at any time within the accuracy of Tycho's observations. Kepler boasted that he would have a solution in eight days. Actually, it took him more than six years before he mastered the problem!

His results, published in 1609, could be stated in two laws that applied not only to Mars but also to any planet. Before we can learn about these laws, we must digress to **Figure 2.2** to see what an "ellipse" is.

Kepler's two laws constitute a complete denial of Plato's doctrine since they assert that the motion of a planet is neither circular nor uniform, but elliptical at variable speed. The two laws are illustrated in **Figure 2.3**.

Ten years later, Kepler published his third law, a mathematical relationship between a planet's average distance to the Sun, and the time it takes to complete one orbit around the Sun. In general terms, the law states that, the greater the average distance of a planet from the Sun, the longer it takes to complete one orbit.

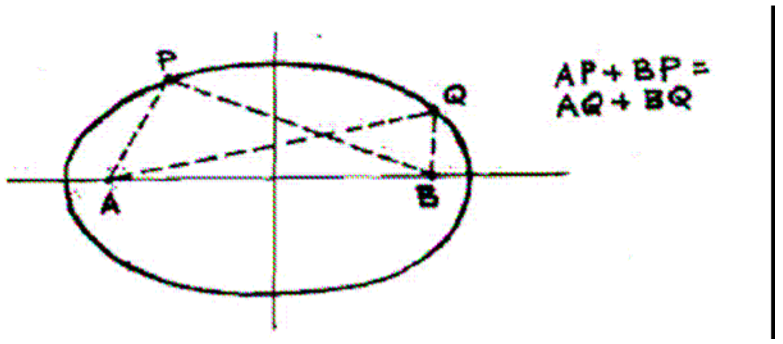

Figure 2.2 - An ellipse

To draw an ellipse, imagine driving two nails on a board at points A and B, and tying to them the ends of a string longer than the distance AB. If you keep the string taut with the point of a pencil and move it around, you will draw an ellipse. Points A and B are each called a "focus" of the ellipse.

By varying the distance between the "foci" and the length of the string, you can vary the size and shape of the ellipse. The closer the foci, the more the ellipse resembles a circle. In fact, when the foci coincide, they become the center of a circle.

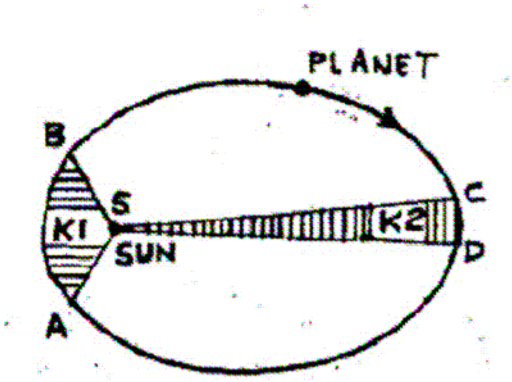

Figure 2.3 - Kepler's 1st and 2nd laws

Both the Sun and a planet are shown above as points. Kepler's first law states that the orbit of a planet is an ellipse with the Sun at one focus.

Kepler's second law states that the line that connects the Sun and a planet sweeps equal areas in equal times, such as the areas K1 and K2 shaded above. This means that the longer arc AB is traversed in the same time as the shorter CD. Thus, within its orbit, the closer a planet is to the Sun, the faster it moves.

Kepler's system superseded the complex systems of epicycles and deferents that had been used before him. It was more than another geometrical scheme contrived to predict planetary positions. It described the actual motions of the planets with respect to the Sun, and did so in mathematical terms.

Still, Kepler lacked an explanation for what made the planets move the way they did. He tried to solve this problem by invoking magnetism, the only force known at the time that appeared to be cosmic in nature. Kepler vaguely speculated that some sort of magnetic force emanating from the Sun acted to push each planet in the direction of its orbital motion.

Chapter 3

LAWS OF MOTION

"If I have seen farther than others,
it has been by standing on the shoulders
of giants."

Isaac Newton

What Kepler lacked to explain the motions of the planets was a knowledge of "mechanics," the new science of motion that emerged from the research of Galileo and Newton.

GALILEO

Galileo Galilei (Italian, 1564-1642) was born twenty-one years after the death of Copernicus. In 1581, he entered the University of Pisa to study medicine, but soon became interested in mathematics. Four years later, lack of money forced him to stop his studies before receiving a degree, but he continued to devote himself to mathematics and natural science. In 1589, a paper he published earned him the title of "Archimedes of his time" and a post as mathematical lecturer at the University of Pisa. Here, he started to conduct experiments on the motion of falling bodies.

At the age of 28, he became professor of mathematics at the University of Padua, where he continued his research on the principles of motion. During the 18 years he spent there, he acquired an international reputation as a scientist and inventor. Although he did not invent the telescope, in 1609 he developed the first telescope suitable for astronomical observation, thus becoming the first person ever to gaze at the skies through a telescope.

With his new instrument, Galileo made a number of important discoveries, such as:

- The surface of the Moon is not smooth and polished, as a perfect celestial body was supposed to be, but rough and uneven.
- There are spots on the surface of the Sun, another supposedly perfect celestial body.
- Four satellites revolve around Jupiter, disproving Aristotle's theory that all celestial bodies revolve naturally around the center of the Universe.

Shortly after announcing his discoveries in 1610, at 46 Galileo was appointed "first philosopher and mathematician" to the Grand Duke of Tuscany, a post that assured him a large salary and more time for research.

For many years, Galileo had believed in Copernicus' theory, and had developed many arguments to support it. Fearing ridicule, however, he had kept his views to himself. The flattering reception he received at the papal court when he demonstrated his telescope encouraged him to defend openly the Copernican theory, and his views became very popular.

Galileo's astronomical discoveries cast serious doubts on Aristotle's System of the World, but they did not actually prove that the Earth revolved around the Sun. Galileo could prove, however, that the arguments raised against the Earth's motion were not valid. His opponents questioned, for instance, how the Earth could appear to be standing still if indeed it was rushing through space. Even a bird perched on a branch, they claimed, would be left behind in space the moment it let go of the branch! Galileo's answer was that the bird was moving with the Earth while it was on the branch, and it retained that motion even after it let go. For the same reason, if a stone is dropped from the mast of a ship, it will hit the deck at the foot of the mast, regardless of whether the ship is standing still or moving, because the stone retains whatever speed it had before it was dropped.

When some Aristotelian scholars started instigating Church authorities against Galileo, the Church did not support them at first. As soon as Galileo, however, tried to maintain that the Copernican theory could be reconciled with the Bible, he met opposition from the theologians. Concerned that the issue might undermine Catholicism in its fight against Protestantism, in 1616 (73 years after Copernicus' death) the Church declared the Copernican theory "false and erroneous". Galileo was admonished not to support the theory, although it could still be discussed as a mere "mathematical supposition."

For the next seven years, Galileo kept silent on the issue. In 1623, Maffeo Barberini - who as a cardinal had been a long-time friend and protector of Galileo - became Pope Urban VIII. Galileo asked his permission to discuss the arguments for the Sun-centered Copernican theory versus the Earth-centered Ptolemaic theory. His request was granted, provided he discussed Copernicus' theory only as a speculative hypothesis, and reached the conclusion that man could not presume to know how the world is really made, since God could have brought about the same effects in unimagined ways.

Nine years later, in 1632 Galileo published his "Dialogue Concerning The Two Chief World Systems - Ptolemaic and Copernican", which quickly won international fame. Galileo's enemies were quick to point out to the

Pope that, in spite of its final conclusion (the one the Pope had prescribed), the book was a compelling defense of the Copernican theory, and seemed to ridicule the Pope.

The following year, even though many people in the Church supported Galileo and would continue to do so privately, Galileo was summoned to Rome, brought to trial and forced to recant. The sentence (which three cardinals refused to sign) would have called for imprisonment, but the Pope immediately commuted it into house arrest, which remained in effect throughout the last eight years of Galileo's life.

Even in the confinement of his small estate, Galileo continued his scientific work to the very end. In 1634, he completed "Dialogue Concerning Two New Sciences". (The new sciences were Strength of Materials and Mechanics). In the second half of the book, he summarized his experiments and thoughts on the principles of mechanics. His last telescopic discovery was made in 1637, only months before he became blind. He died in 1642 at the age of 78.

Almost three and a half centuries later, in October of 1992, Pope John Paul II made a speech vindicating Galileo, in which he said "In the 17th century, theologians failed to distinguish between belief in the Bible and interpretation of it. Galileo contended that the Scriptures cannot err but are often misunderstood. This insight made the scientist a wiser theologian than his Vatican accusers". [1]

Some Basic Definitions

Before summarizing the results of Galileo's research on motion, we need a few basic definitions and concepts.

Speed

Let us imagine ourselves driving along a straight road. If it takes us 2 hours to cover a distance of 100 miles, we say that we have traveled at the average speed of 50 miles per hour. Most likely, we have moved at times faster and at times slower than this average speed. Had we watched our odometer at half-hour intervals, we might have found that, in the first half-hour, we covered a distance of 15 miles, therefore moving at an average speed of 30 miles per hour. Again, within this half-hour interval, we may have moved sometimes faster and sometimes slower than the average speed for this interval. By considering average speeds for smaller and smaller intervals of time, we arrive at the concept of instantaneous (instant-by-instant) speed.

We say that a body (object) moves with straight uniform motion during some given interval of time, if both the speed and the direction of motion remain constant throughout this interval.

Acceleration

Let us assume now that a body moves, still along a straight line, at a speed that is not uniform. Let us assume, for instance, that the speed is 30 feet/second at some point in time, and it is 50 feet/second two seconds later. We say that the body has "accelerated" from 30 to 50 feet/second in 2 seconds, and we define the average acceleration during this time interval as the change in speed divided by the time interval, or (50-30)/2 = 10, that is, the speed has increased by an average of 10 feet/second each second. Conversely, if the speed of a body *decreases* from 50 to 30 feet/second in two seconds, we say that the body has "decelerated" at the average rate of 10 feet/second each second.

As we did for speed, we can consider smaller and smaller time intervals to arrive at the concept of instantaneous acceleration. In general, both the speed and the acceleration of a moving body may change from instant to instant.

We say that a motion along a straight line is uniformly accelerated during some interval of time, if the acceleration is constant during this interval; for instance, if its speed increases steadily at the rate of two feet per second every second.

Vectors

In general, an object will move along some curved path, rather than a straight line. Speed alone is thus not sufficient in general to describe the motion of a body, since we must also specify its direction, which may vary. For this reason, it is convenient to introduce a new quantity called "velocity", which specifies both the speed and the direction of a moving object.

Velocity is an example of a quantity that can be represented by an arrow that has a certain length, or "magnitude", and a certain direction. We call such an arrow a "vector". For a given moving object, at some instant, the direction of its velocity vector gives the direction of motion, while the length of the vector represents the speed of the object.

Another example of a quantity that can be represented by a vector is the *force* with which we might pull a boat at the end of a rope. The direction of the force vector is the direction of the taut rope; the length of the vector represents the strength with which we are pulling the boat. Another example of a force is what acts on a small piece of iron, when attracted by a nearby magnet.

Given two vectors, it is possible to replace them with a single equivalent vector, see **Figure 3.1**.

Galileo's results

Galileo proved wrong two major contentions of Aristotle's that had gone undisputed for many centuries:

- Heavy bodies by their nature fall faster than light bodies.
- The natural state of a body is to remain at rest. To put a body in motion and keep it moving requires the action of a force constantly in touch with the body.

To study motion, Galileo used small, smooth metal spheres of different weights rolling down a very smooth inclined plane. The use of small spheres and the smoothness of the plane greatly reduced the effects of air resistance and friction.

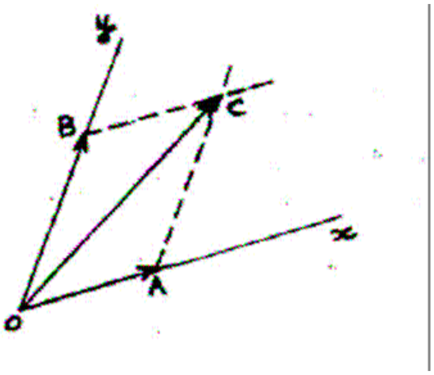

Figure 3.1 – Adding two vectors

Given two vectors OA and OB, we determine, as shown, their “sum” or “resultant,” the vector OC. (Lines OA and BC are parallel, and so are OB and AC.) A boat pulled by two forces represented by OA and OB will move as if it were pulled by a single force represented by OC. Conversely, given the vector OC and the directions x and y, we can determine the “components” of OC along the directions x and y.

To find the “difference” of two vectors, we simply add to the first the “opposite” of the other (same length but opposite direction).

A sphere rolling down an inclined plane is not affected by its full weight W (see **Figure 3.2**) but only by the component of W that is parallel to the inclined plane. The greater the slope of the plane, the greater is the component of W parallel to the plane. Thus, by changing the slope of the plane, Galileo could control how quickly, or how slowly, the spheres rolled down.

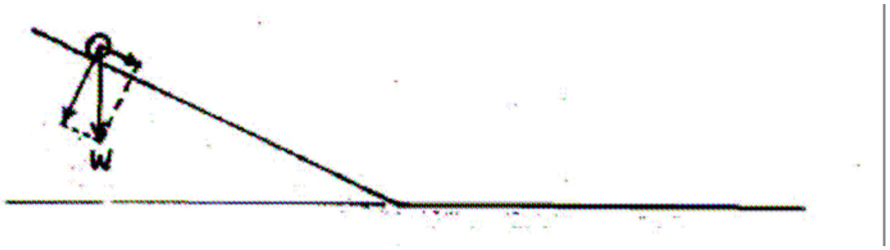

Figure 3.2 – Galileo’s inclined plane

Galileo's major findings and conclusions were:

- A sphere rolls down with constant acceleration. If, for instance, the sphere travels one inch during the 1st second, it will travel 3in. during the 2nd second, 5in. during the 3rd, 7in. during the 4th, and so on (thus, the acceleration is 2in. per second every second).
- For a given slope of the inclined plane, the acceleration does not depend on the weight of the rolling sphere.
- If we increase the slope (and thus the force acting on the sphere), the sphere will roll down with a higher constant acceleration. Galileo concluded from his experiments that, in a vacuum (i.e., in the absence of any resistance), regardless of weight, a body would fall with a constant acceleration "g" equal to about 32 feet/second every second.
- When the rolling sphere reaches the bottom of the incline, and moves on to a smooth horizontal plane, its speed remains constant, at least for a while. The sphere eventually stops because the horizontal surface is not perfectly smooth, and the resistance encountered by the sphere causes it to decelerate. If the smoothness is increased, the sphere will roll over a longer distance.

Galileo concluded that, if there were no decelerating forces, the sphere would continue to roll forever at constant speed. Contrary to what Aristotle had claimed, no force was needed to keep the sphere moving at constant speed.

Galileo studied also the motion of a cannon ball shot with some initial velocity V at an angle to the horizontal, see **Figure 3.3**. He assumed the motion to occur in a vacuum, with no forces present except for gravity. In the horizontal direction, the cannon ball travels at a constant speed, since there are no decelerating forces in this direction. In the vertical direction, instead, the cannon ball will be steadily decelerated by gravity until, having reached some height, it will come to a halt, and will then start falling with constant acceleration. Galileo could show mathematically that the trajectory of the cannon ball would be a "parabola", a "conic" curve well familiar to the Greeks.

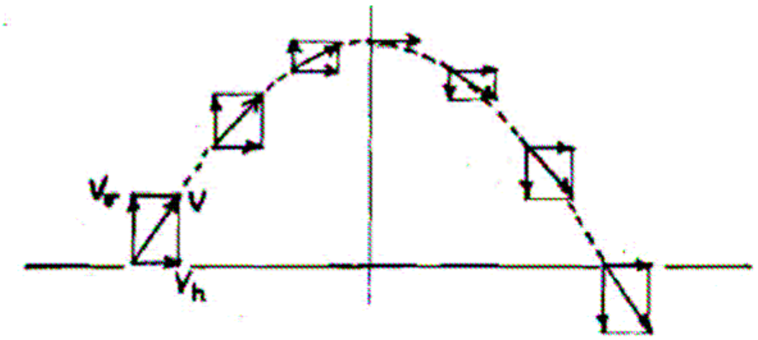

Figure 3.3 - Trajectory of a cannon ball

By concentrating on how motion would occur under idealized conditions, i.e., in the absence of any resistance, Galileo was able to understand the fundamental nature of motion better than his predecessors had been. The presence of a force that is in line with the motion of a body will cause the body to accelerate or decelerate, depending on the direction of the force relative to the motion. The absence of any forces will result in zero acceleration, that is, constant velocity. If the body is at rest, it will remain at rest; if the body is moving at some speed, it will continue to move at that speed in the same direction.

Galileo was the first to bring to the study of physical phenomena a method of investigation that combined reasoning, experimentation and mathematics. He created the modern idea of the experiment, in which special conditions are created to verify theoretical deductions.

About the role of mathematics in nature, Galileo wrote:

"Philosophy is written in this grand book, the universe, which stands ever

open to our gaze. But the book cannot be understood unless one learns to comprehend the language and read the letters in which it is composed. It is written in the language of mathematics, and its characters are triangles, circles and other geometric figures without which it is humanly impossible to understand a single word of it; without these, one wanders about in a dark labyrinth."

NEWTON

The same year Galileo died, Isaac Newton (1642-1727) was born in a small village in central England. His father, a farmer, had died three months before. Two years later, his mother married a prosperous minister from a nearby village and moved there to raise three stepchildren, leaving her son in her mother's care. For nine years, until his stepfather died, young Isaac was separated from his mother. This early traumatic experience may have been the cause for the strong sense of insecurity that afflicted Newton throughout his life. Reluctant to face controversy and criticism, he would become much more interested in the private pursuit of science than in making his results known to others.

In 1661, he entered Trinity College in Cambridge. One of his professors, Isaac Barrow, encouraged him to study mathematics and optics (the study of phenomena pertaining to light). In less than a year, Newton mastered on his own the literature of mathematics, and started some original work. He kept to himself, however, the results of his private research.

In 1665, he received his bachelor's degree. That same year, when the Great Plague spread to Cambridge, the college was closed, and Newton returned to his native village. Here, during two extraordinary years, he laid the foundations of his monumental contributions to mathematics, optics and mechanics:

- He invented a new system of mathematical analysis called calculus.
- Experimenting with glass prisms, he discovered that a narrow beam of sun light, after passing through a prism, generates on a screen a series of colored bands in the familiar pattern of the rainbow: red, orange, yellow, green, blue, indigo and violet. He concluded that white light consists of rays of different colors, which are bent to different extents when passing through a prism, thus creating the rainbow spectrum on the screen.
- In mechanics, having analyzed uniform circular motion, he applied his analysis to the motion of the Moon and the planets, and derived the "inverse square law". This law states that, as a planet moves along its orbit, a force is acting on it that is inversely proportional to the square of

the planet's distance to the Sun[3], and is directed toward the Sun.

In 1667, after the university reopened, Newton returned to Trinity College. Two years later, at the age of 27, he was appointed professor of mathematics, succeeding Barrow, who recommended him as a man of "unparalleled genius".

In 1671, the Royal Society, England's prestigious academy of science, invited Newton to demonstrate a new type of telescope he had developed. Encouraged by the enthusiastic reception given to him, Newton volunteered to present a paper on his theory of colors early in 1672. The paper was well received but not without some dissent.

This led to the first of a number of bitter controversies that troubled Newton throughout his life, because he was incapable of dealing rationally with criticism. Less than a year later, he withdrew into virtual isolation, which lasted some twelve years. During these years, he immersed himself in alchemy.

By the mid-1680s, although he had already made his major discoveries in calculus, optics and mechanics, Newton had published very little. In the meantime, a number of scientists were trying to define a force originating from the Sun that could account for the planetary motions described by Kepler.

Among these scientists was Edmond Halley, the British astronomer after whom the famous comet was named. Like Newton, Halley and others had concluded that the force that keeps a planet in orbit must decrease as the square of the planet's distance from the Sun. They had not been able, however, to derive mathematically the orbit that would result from such a force.

In 1684, Halley visited Newton about this problem, and was astonished to find that Newton already had the solution. Newton's problem was that he had mislaid his calculations to prove it! Halley made him promise to send the proof, which arrived three months later.

In about 18 months between 1685 and 1686, Newton wrote in Latin his masterpiece "Philosophiae Naturalis Principia Mathematica" (Mathematical Principles of Natural Philosophy). The book, written in the style of Euclid's Elements, used intricate geometric arguments. It was published in 1687.

In the Principia, Newton was able to show how the behavior of objects falling near the surface of the Earth, the tides, and the motions of the celestial bodies could all be mathematically predicted from four laws he

---

[3]. *When the distance to the Sun doubles, the force becomes 4 (2x2) times smaller; when the distance triples, the force becomes 9 (3x3) times smaller, and so on.*

postulated. Contrary to Aristotle's fundamental distinction between heavens and Earth, both followed the same mathematical laws! There were no precedents for such elegant simplicity and universality. His masterpiece immediately raised him to international fame, which was followed by many great honors.

In 1699, Newton became Master of the Mint (responsible for the coinage of money), a lucrative position that he held until his death. In 1703, he was elected president of the Royal Society. In 1705, he was knighted, becoming the first scientist ever to be so honored.

In 1704, Newton finally published his book "Optiks", which contained mostly work he had done some 30 years before on the study of light. His work on calculus, which was even older, was included merely as an appendix. In the meantime, calculus had been independently invented by the German mathematician and philosopher Leibniz, who published his work 20 years before Newton. A long and bitter dispute followed, with charges of plagiarism raised by supporters on both sides.

These accusations of dishonesty unleashed Newton's temper. He wrote papers in his defense and published them under the names of obliging supporters. As president of the Royal Society, he appointed an "impartial" committee to investigate the issue, and secretly wrote its report.

Yet, Newton was also capable of charming modesty, as suggested by the following remark, made shortly before his death:

> *"I do not know what I may appear to the world, but to myself*
> *I seem to have been only like a boy playing on the seashore,*
> *and diverting myself in now and then finding a smoother*
> *pebble or a prettier shell than ordinary, whilst the great*
> *ocean of truth lay all undiscovered before me."*

When Sir Isaac Newton died in 1727 at the age of 85, the greatest honors, including a royal funeral, were accorded to him.

The Laws of Motion

At the beginning of the Principia, in Euclidean fashion, Newton gives some definitions and then offers three laws of motion as unproven "axioms": the first two restate and generalize Galileo's findings, the third introduces a new and crucial concept.

Newton did not claim that his axioms or "laws" were self-evident. He presented them as working assumptions to be accepted only as long as they helped to explain in exact detail previously unexplained phenomena.

Before we can present these laws of motion, we need to broaden the concept of acceleration previously introduced for motion on a straight line. In the most general case, as a body moves along some curved path, its velocity vector may change from point to point in direction and/or speed.

The average acceleration from point X to point Y must now be defined as the *vector* difference of the two velocities at X and Y divided by the time it takes to go from X to Y. As we did before, by considering points X and Y that are closer and closer, we can arrive at the concept of an instantaneous acceleration vector. For our purposes, it will suffice to keep in mind that

- Like velocity, acceleration is a vector.
- Acceleration occurs whenever there is a change in speed and/or direction.

Another key concept is that of "mass". It was defined by Newton as the "quantity of matter" of an object, and was assumed to be constant. It is a purely numerical quantity, not a vector.

The mass of an object should not be confused with its weight: it is now well known, for instance, that the body of an astronaut weighs less on the Moon than on Earth, although it has the same mass in either place.

Law I (the law of inertia)

If a body is at rest or moving at constant speed in a straight line, it will remain at rest, or will continue to move at constant speed in a straight line, as long as no force is acting on it.

Passengers inside a car that is suddenly slowed down or stopped find themselves continuing in their forward motion, in accordance with the first law. They might even go through the windshield, if not restrained by a seat belt.

If the car turns suddenly and sharply, its passengers tend to continue in their original direction and slide across the seat. Thus, a body tends to maintain both its speed and its direction of motion.

Law II (the law of force)

If we wish a body of mass m to acquire an acceleration that has the value A in some direction, we must apply to it a force which has the same direction as the acceleration, and a value F equal to the mass m times the acceleration A

$$F = m \times A$$

The second law tells us that, the larger the mass of a body we want to accelerate, and the larger the acceleration we want it to attain, the larger must be the force we need to apply to the body.

For a given force, the larger the mass of the body to which the force is applied, the smaller the acceleration that will result; the smaller the mass, the larger the acceleration.

To maintain a constant speed (zero acceleration), no force is required, regardless of the speed. To maintain a steady acceleration, a constant force is required. To increase the acceleration, the force must be increased.

Law III (the law of action and reaction)

To every action, there is always an equal and opposite reaction, i.e., if P exerts a force on Q, then Q exerts an equal and opposite force on P.

What happens when we punch a brick wall provides a painful illustration of the third law! Another example is the recoil of a cannon as it propels an iron ball.

The Motion of a Body

Galileo's work showed that, if a body is already moving with some velocity, and a constant force is applied to it *in the direction of its motion*, the body will start accelerating at a constant rate in the same direction. If the force, instead, is in the *opposite* direction, the body will start decelerating at a constant rate.

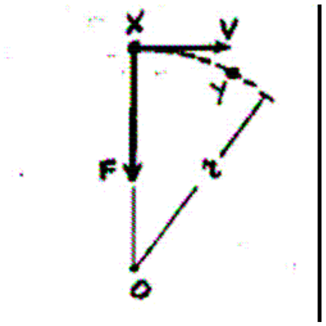

Figure 3.4 - Force and velocity vectors are perpendicular to one another

On the other hand, Newton analyzed what happens if a body is already moving with some velocity V, and a force F is applied to it *at 90 degrees* to the direction of motion as in **Figure 3.4**.

In the next instant, the speed of the body will not change, but its direction will be "bent" along a very short arc of a circle. The smaller is the radius of this circle, the larger is the bending. For a body with a given mass,

- the larger the velovity, the smaller the bending.
- the larger the force, the larger the bending.

Thus, we might say, there is a struggle going on between the velocity of a body and a perpendicular force applied to it: the velocity opposes bending, the force wants to create it.

In the most general case, the force vector applied to a body will have a component in the line of motion - which will either increase or decrease the speed - and a component perpendicular to the line of motion, which will cause a "bending" in direction.

If we know the position and velocity of a body at some particular point, and we know the "law" by which the force vector acting on the body changes in magnitude and direction from point to point, we can determine the subsequent path of the body.

The Law of Universal Gravitation

Newton made the very bold assumption that the attraction between the Sun and a planet was just a special case of a much more general situation: any two bodies in the universe with masses m1 and m2 exert on each other a force of attraction along the line connecting them. The magnitude of this force is

- directly proportional to the product of the two masses (i.e., if the product doubles, the force doubles; if the product triples, the force triples, and so on), and
- as previously mentioned, inversely proportional to the square of the distance.

Qualitatively, the law states that the larger the two masses, the larger the force of gravity between them; the larger the distance, the smaller the force.

Orbital paths of planets and comets

With his theory of gravity, Newton could account for Galileo's experimental results. He could also derive mathematically the three laws that Kepler, after years of painstaking labor, had obtained empirically from Tycho Brahe's data.

To understand in a general way the motion of a planet around the Sun along an elliptical orbit, we have to consider the interplay at any instant between the velocity of the planet and the gravitational pull of the Sun, which acts as a gigantic rubber band. In one half of the orbit, the "rubber band" *slows down* the planet while bending its path, until at some point, the planet starts coming around. In the other half of the orbit, the "rubber band" *speeds up* the planet while bending its path, until at some point the planet starts turning around, and the cycle repeats; see **Figure 3.5**.

The interplay between force and velocity explains also how orbits of different shapes and sizes are possible. If a comet, for instance, approaches the solar system with a sufficiently high velocity, its path will be deflected somewhat, but the comet will continue on its cosmic journey, never to return. (The "rubber band" never succeeds in turning the comet around.) On the other hand, if the approaching comet does not have sufficient velocity, it will be "captured" by the gravity pull of the Sun, and forced to orbit along an ellipse more or less elongated, depending on what its approach velocity was. The famous Halley's Comet, for instance, traveling at a speed of more than 80,000 miles per hour, passes by the Earth and the Sun at intervals of about 76 years.

In addition to the main interactions between the Sun and each planet, there are smaller gravitational interactions between planet and planet. Since the Sun accounts for more than 99% of the mass of the entire solar system, its effect on the planets is dominant, but not exclusive.

Perhaps the most dramatic confirmation of Newton's theory was provided by the discovery of the planet Neptune. In 1781, Sir William Hershel accidentally discovered a planet, which was later called Uranus. The orbit of the new planet, however, was not in agreement with Newton's theory, even taking into account the gravitational effects of the two large planets nearby, Jupiter and Saturn. If Newton's theory was valid, there had to be an undiscovered planet to account for the irregularities of Uranus.

The orbit of this hypothetical planet was computed by the French astronomer Le Verrier. On September 23, 1846, after only one hour of searching, the German astronomer Galle found the eighth planet, Neptune, at the precise spot in the sky that had been suggested to him by Le Verrier. (The ninth and last planet, Pluto, was accidentally discovered in 1930.)

The World as a Machine

The success of Newton's theory was so dramatic that he would dominate Western thinking almost as much as Aristotle had done for many centuries before. Eventually, Newtonian mechanics suggested to some that, if the positions and velocities of all particles of matter could be known at one particular instant, it would be possible - in principle at least - to determine the entire past and the entire future of every particle, and therefore the entire course of the universe.

In this view, the universe has the ordered structure, not of an Aristotelian organism, but of a machine, a gigantic clock that, once wound, proceeds on its own in predetermined, unalterable mechanical fashion.

This was not, however, the worldview held by Newton, who believed that God would have to intervene from time to time to maintain the stability of the solar system. A century later, the French mathematician, astronomer and physicist Pierre-Simon LaPlace proved that the solar system would remain stable on its own. Supposedly, when Napoleon asked him "What about God?", LaPlace replied "I did not need that hypothesis."

In his final remarks at the end of the Principia, Newton wrote: "it is not to be conceived that mere mechanical causes could give birth to so many regular motions. .... This most beautiful system of the sun, planets and comets could only proceed from the counsel and dominion of an intelligent and powerful Being. ..... All that diversity of natural things ... could arise from nothing but the ideas and will of a Being necessarily existing. .... Thus much concerning God; to discourse of whom from the appearances of things, does certainly belong to natural philosophy."

There was one aspect of the new theory that was disturbing even to Newton: *instantaneous action at a distance*. As he admitted in a letter, "that one body may act upon another at a distance through a vacuum without the mediation of anything else, ... is to me such an absurdity that I believe no man who has in philosophical matters a competent faculty of thinking can ever fall into it".

Newton did not claim to have an explanation for gravity. Toward the very end of the Principia, he remarked: "Hitherto I have not been able to discover the cause of those properties of gravity from phenomena, and I feign no hypothesis. .....To us it is enough that gravity does really exist, and act according to the laws which we have explained, and abundantly serves to account for all the motions of the celestial bodies, and of our sea."

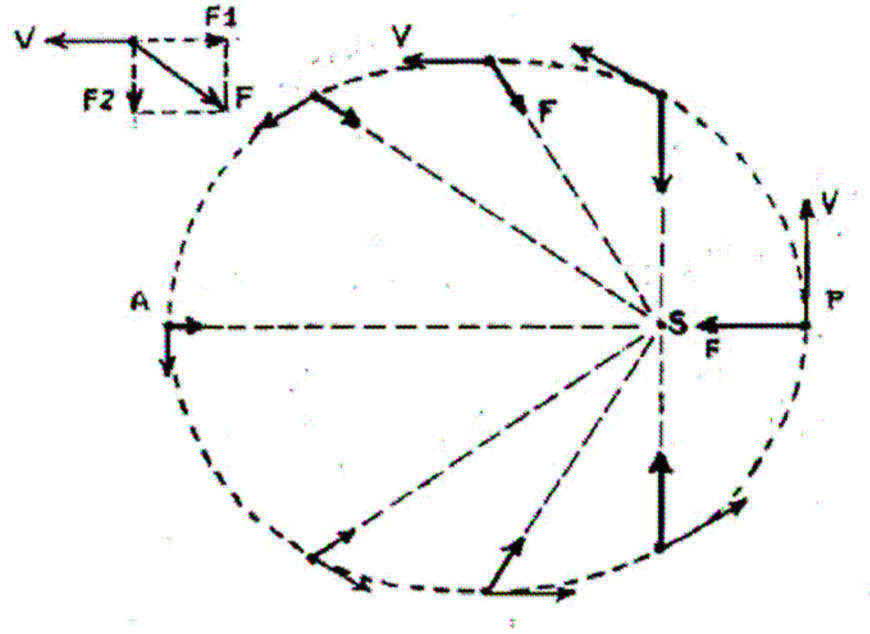

Figure 3.5 - A planet's elliptical orbit

Point S represents the Sun fixed at one focus of the ellipse. At any point on the orbit, the force vector F, which represents the pull of gravity, is always pointed toward S. Because of the inverse square law, its magnitude decreases with the planet's distance from point S. In the upper half of the orbit, from point P to point A, the angle between the force and velocity vectors is always more than 90 degrees, that is, the component of the force in the line of motion opposes the motion. As a result, the force causes the velocity to decrease in value, while at the same time forcing the path to bend.

In the lower half of the orbit, instead, the angle between the force and velocity vectors is always less than 90 degrees, that is, the component of the force in the line of motion aids the motion. As a result, the force causes the velocity vector to increase in value from A to P, while at the same time forcing the path to bend.

Starting from point P, we might say that the planet, under the impetus of its velocity, tries to escape the gravitational pull of the Sun, which keeps slowing down the planet and bending its path, until at point A, the planet starts coming around. The gravitational force now starts pulling the planet toward the Sun, increasing its speed and bending its path, until the planet is brought back to P, to start all over again.

Newton indeed had abundantly solved the riddle of the heavens, which had puzzled the human mind for thousands of years. Actually, we should say, he had found a solution. Less than two and a half centuries later, Albert Einstein would arrive at a totally different conception of gravity.

Chapter 4

ENERGY

If we reflect on the preceding two chapters, two observations come quickly to mind. First, they cover a considerable span of time: almost two thousand years elapsed from Euclid's Elements to Newton's Principia, published in 1687. Second, they are primarily concerned with the study of motion, particularly the motion of celestial bodies.

In the wake of Newton's stunning triumph, science gained great new impetus. Its pace accelerated and its scope broadened considerably. In a span of little more than three centuries, fundamental aspects of energy, matter and light were explored with great success and whole new fields of investigation were started.

The concept of "energy" is one of the most basic in all of physics. But, what is energy? We might think of it as the currency used to conduct nature's transactions. In any process of nature, we might say, there is always some amount of energy that "changes hands" to "pay" for something.

A certain amount of wealth, initially in dollars, might later be partially converted into German marks, part of which might later still be converted into Japanese yens. In the absence of any losses (or commissions), whatever combination of dollars, marks and yens we may have, it will still represent the original amount of wealth. Similarly, energy may be converted from one form to another, but the total amount remains constant.

Physicists say that, if something, or somebody, performs "work" on a "system", the system gains "energy" of some form. Later, the system may spend some or all of its energy in the performance of work. When we wind a mantle clock, for instance, the work we do goes into energy that is stored in the clock's spring. As the spring slowly unwinds, its energy is spent doing "work" by turning the gears of the clock mechanism.

MECHANICAL ENERGY

Newton's mechanics led to the definition of two forms of energy called gravitational potential energy and kinetic energy.

Suppose you have been hired to hoist a heavy rock, using a rope and pulley. You certainly would want to be paid in proportion to the weight of the rock and the height to which you hoist it: this is how you would measure your work, and so would a physicist.

Having hoisted the rock to some height, you tie your end of the rope to a hook on the ground, and then you ask: Just what kind of energy has the rock acquired because of my work? It is a latent, potential kind of energy, similar to that of a compressed coil spring, ready, but not yet free, to snap. For our rock, it is called *gravitational potential energy*, and is defined as the product of the rock's weight times its height relative to some reference level.

If you now cut the rock loose, you will see that indeed we have energy there. The rock starts falling, accelerating as it goes. Its potential energy is being converted into another kind of energy called *kinetic energy*, defined as half the mass times the square of the speed ("kinetic" derives from the Greek word for motion). By the time the rock reaches the ground, essentially all the energy you put into it has been converted into kinetic energy. This energy can be used to do useful work, if the initial intent was to crunch something on the ground.

In summary, then, an object has (gravitational) potential energy by virtue of the fact that it has weight at some height; it has kinetic energy by virtue of the fact that it has mass moving at some speed.

Let us consider now a swing in a park. If you give it one good push, it will start oscillating. Having reached a high point, it will swing down, converting potential energy into kinetic energy, as it loses height, while gaining speed. When it reaches its low point, it will swing up, converting kinetic energy into potential energy, as it gains height, while losing speed.

Under ideal conditions of no friction and no air resistance, the oscillations of the swing would go on forever. The two forms of mechanical energy would continuously convert into one another without loss, and their sum would remain forever constant.

Under real conditions, however, there will be only a few oscillations, after which the swing comes to rest. You did some work when you gave that push, and you put some energy into the "system". Where has that energy gone? Because of friction and air resistance, heat has been generated. One of the major triumphs of 19th-century physics was the experimental proof that heat is another form of energy to be entered in nature's balance sheet.

HEAT AS ENERGY

During the 18th century and half of the 19th century, heat was thought of as a weightless fluid, called *caloric*, which could not be either created or destroyed. The notion of heat as a fluid provides a convenient mental image: we talk of heat flowing from a higher to a lower temperature, just as we talk of water flowing from a higher to a lower level.

It was later shown that heat could be generated in any amounts at the expense of mechanical work. If, by means of a crank, for instance, you keep turning a paddle wheel inside some container of water, you will generate heat and warm up the water. In the 1840's, an English physicist (and brewer), James Prescott Joule, concluded after extensive measurements that a certain amount of mechanical work consistently produces a corresponding amount of heat, which must thus be viewed as another form of energy.

Measuring temperature

Closely associated with the notion of heat is that of temperature. A familiar way of measuring temperature is by means of a mercury thermometer, using either the Fahrenheit scale, or the "centigrade" (100-degree) scale. In the latter, 0° C is assigned to the level of the mercury when the thermometer is in a bath of melting ice, and 100° C is assigned to the level of the mercury when in a bath of boiling water. The interval in between is divided into 100 equal spaces. The two temperatures, 0° C and 100° C, correspond to 32° F and 212° F, respectively, on the Fahrenheit scale. In either scale, the zero point has been arbitrarily selected, and does not represent the zero value of anything.

The international standard for scientific measurements of temperature is the Kelvin scale, named after the British physicist Lord Kelvin. In this scale, the zero point is "absolute zero", the lowest possible temperature in the universe. The Kelvin scale for absolute temperatures is similar to the centigrade scale except that -273° C becomes 0° K and, therefore, 0° C becomes 273° K.

THERMODYNAMICS

In the early 1800s, the Industrial Revolution, which originated in Britain in the mid-1700s, was in full swing. It was driven by the steam engine, the first practical device to generate mechanical energy from heat. This was the setting for the birth of a new science, called "thermodynamics." It is concerned with how heat, temperature, work and energy are related. The two most important principles of this science are known as the first and second laws of thermodynamics.

The first law of thermodynamics

In 1847, the German physicist Hermann Helmholtz proposed that the definition of energy be generalized to include forms of energy other than mechanical energy in its kinetic and potential forms.

When work is done on a system, it increases the total energy of that system. The additional energy might go into mechanical energy, heat, electrical energy, or some other form of energy. The system, on the other hand, can expend some of its energy to do work. A system is said to be isolated, when it neither receives energy from, nor gives energy to, its surroundings: its total energy in its various forms must then remain constant.

As Helmholtz stated, "... the quantity of energy in nature is just as eternal and unalterable as the quantity of matter". This Principle of Energy Conservation constitutes the first law of thermodynamics.

The Second Law of Thermodynamics

While nature does not restrict the conversion of work into heat, it does impose severe restrictions on the conversion of heat into work. Basically, the second law of thermodynamics states that such restrictions exist.

Consider what happens in a steam engine. Inside a cylinder, the expansion of hot steam under pressure causes the motion of a piston, which in turn causes the wheels of a locomotive to turn. Part of the steam's heat is thus converted into work; the rest is discharged into the surroundings at some lower temperature.

The heat in the exhaust still constitutes energy, but a *degraded* form of energy. It too can be converted into work, but only if we can make it drop to an even lower temperature. Thus, although energy is conserved in an isolated system, the amount of "free energy", i.e., the energy available to do work, keeps on decreasing. This is one of several ways in which the second law of thermodynamics can be stated.

If the universe as a whole is an isolated system, the second law predicts an inevitable doom. In some very distant future, the universe - totally devoid of "free energy" - will have run down completely. The so-called "heat death" will have occurred. All that will be left is a sea of disorganized matter close to absolute zero temperature.

There is a third law of thermodynamics. It states that the absolute zero temperature is unattainable: as we get closer and closer to this minimum temperature, it becomes more and more difficult to extract additional energy and reduce further the temperature.

Chapter 5

ATOMS AND MOLECULES

As mentioned earlier, Greek philosophy produced two theories on the nature of matter: a theory of fundamental *elements*, and a theory of indivisible *atoms*. Both concepts were eventually incorporated in the scientific theory of matter that evolved during the 1700s and 1800s.

In the Middle Ages, alchemists engaged in occult practices, whose main goals were to transform base metals such as lead and copper into gold and silver, and to discover an elixir of perpetual youth.

Modern chemistry gradually evolved from medieval alchemy and early medicine into a quantitative science. Among its early accomplishments was the distinction between "chemical compounds" and "elements". Most forms of matter are compounds, or mixtures of compounds. In order to find the elements that form a compound, the compound is chemically broken down into its components; these components, in turn, are broken down into their components, and so on, until no further breakdown is possible. The substances that cannot be broken down further are, by definition, "elements".

By this process, in the 19th century, it was established that such substances as hydrogen, oxygen, copper, lead, silver and gold are elements. The present count of chemical elements is somewhat over 100, a relatively small number compared to the very large number of compounds observed in nature.

Dalton's Atomic Theory

On the basis of extensive experimental data, the British chemist and physicist John Dalton (1766-1844) formulated a "*law of definite proportions*". It states that, for any chemical compound, the weights of its ingredients are always in fixed proportions. For instance, when a sample of water is broken down and its two constituents, hydrogen and oxygen, are weighed, their ratio is invariably 2 to 16, in round numbers, regardless of the size of the sample. For carbon monoxide, the ratio of carbon to oxygen is 12 to 16.

To account for this, Dalton proposed an atomic theory of matter, which was published in 1808. He speculated that a sample of any element cannot be subdivided indefinitely into ever smaller pieces, but consists of some very large number of identical indivisible atoms. Atoms of different elements have different weights. Thus, there are as many kinds of atoms as there are elements. The most basic unit of any compound of two or more elements is called a "molecule", in which two or more atoms are bonded together. In a chemical reaction, the atoms remain unchanged but become bonded in different combinations.

If a sample of a compound consists of a very large number of identical molecules, and, within each molecule, the atoms of the elements involved are in some fixed ratio by weight (say, 1 to 8), the same ratio must apply to the sample as a whole, which is what the law of definite proportions states.

Atomic Weights

Eventually, it was determined how many atoms of what elements make up the molecules of the compounds they form. Dalton's law could be used then to determine the *relative* atomic weights of the various elements.

The table below lists a few familiar elements and their *relative* atomic weights, in round numbers. Hydrogen is the lightest of all atoms, followed by helium, which is about 4 times heavier. Note the substantial difference in atomic weights between hydrogen and uranium.

Hydrogen	1
Helium	4
Oxygen	16
Aluminum	27
Iron	56
Silver	108
Gold	197
Uranium	238

The Periodic Table

In 1869, the Russian chemist Dmitry Mendeleyev proposed his "periodic law", whereby the chemical elements show a periodic recurrence of certain physical properties, when they are arranged in the order of increasing atomic weights.

Mendeleyev introduced the first "periodic table", in which the 63 elements then known were arranged in rows and columns, with gaps here and there. An abbreviated version of the current periodic table appears in Table I, which shows the names of most elements and, in some cases, their standard symbols. The current periodic table accounts without gaps for all the elements presently known, including 12 that have been synthetically prepared. The known elements are identified by a sequential number ranging from 1 to 106. (The significance of this number will be explained later.)

Table I
PERIODIC TABLE OF ELEMENTS

1 H hydrogen								2 He helium
3 Li lithium	4 Be beryllium		5 B boron	6 C carbon	7 N nitrogen	8 O oxygen	9 F fluorine	10 Ne neon
11 Na sodium	12 Mg magnesium		13 Al aluminum	14 Si silicon	15 P phosphorus	16 S sulphur	17 Cl chlorine	18 Ar argon
19 K potassium	20 Ca calcium	a	31 Ga gallium	32 Ge germanium	33 As arsenic	34 Se selenium	35 Br bromine	36 Kr krypton
37 Rb rubidium	38 Sr strontium	b	49 In indium	50 Sn tin	51 Sb antimony	52 Te tellurium	53 I iodine	54 Xe xenon
55 Cs cesium	56 Ba barium	c	81 Tl thallium	82 Pb lead	83 Bi bismuth	84 Po polonium	85 At astatine	86 Rn radon
87 Fr francium	88 Ra radium	d	113	114	115	116	117	118

a	10 elements (#21 to #30), including chromium (#24), manganese (#25), iron (#26), cobalt (#27), nickel (#28), copper (#29) and zinc (#30)
b	10 elements (#39 to #48), including silver (#47)
c	24 elements (#57 to #80), including platinum (#78), gold (#79), and mercury (#80)
d	18 elements (#89 to #106), including uranium (#92) and plutonium (#94)

The periodic table is an intriguing blueprint of matter. It has 7 rows containing a variable number of elements (2, 8, 8, 18, 18, 32 and 20, for a total of 106). Elements that appear in the same column all display similar properties. For instance, the 6 elements in the last column on the right, helium (#2) to radon (#86), called the "rare gases", are all odorless, tasteless, colorless, nonflammable gases with very limited, if any, tendency to combine with other elements. The periodic recurrence of physical properties of elements remained unexplained until the late 1920's.

Hydrogen, the first element in the periodic table, is by far the most abundant in the universe. By mass, it accounts for about 75% of the visible matter in the universe. Helium, the second element, accounts for about 23%; all the remaining elements together account for a mere 2%.

Chapter 6

PARTICLES IN MOTION

From the 17th century on, physicists were aided in their interpretation of nature by a "kinetic theory" of matter as consisting of very small particles, or molecules, in very rapid random motion. With this theory, external properties of matter could be explained by the internal behavior of particles in motion.

The easiest illustration is provided by gases, which have two major properties: 1) they tend to expand until they completely fill any container into which they are introduced, and 2) they are easily compressed under normal conditions, and have a density about 1000 times less than liquids.

These properties are readily explained in terms of the kinetic theory. A gas tends to fill its container because it is made of a huge number of molecules free to move very rapidly at random in all directions. A gas can be easily compressed because, normally, its molecules are very widely separated by empty space.

At the other extreme, the molecules of a solid are strongly bonded together by cohesive forces, which, as we will see, are electrical in nature. These cohesive forces are not as strong in the case of liquids, which, accordingly, have properties that are intermediate between gases and solids. The molecules of a liquid are almost in contact with one another, but are free to slide over one another. Consequently, a liquid is hard to compress, but free to assume the shape of its container.

Molecules have limited mobility in liquids, and essentially none in solids. At all times, however, there is motion *within* each molecule, with its constituent atoms vibrating back and forth. Typical vibrations are in the order of 10 million millions per second!

The Empirical Study of Gases

Important early clues about the nature of matter were provided by the empirical study of gases, which are much easier to investigate than liquids or solids.

In the 17th century, the Anglo-Irish chemist Robert Boyle experimentally studied the behavior of a gas contained in a cylinder with a movable piston, see **Figure 6.1**.

Later, physicists established empirically a general law, called the "ideal gas law". Provided the pressure is not too high nor the temperature too low, this law states that

> the product of (the pressure exerted by the gas) times
> (the volume of the gas)
> is proportional to (the absolute temperature of the gas).

Statistical Mechanics

The rather simple external behavior described by the ideal gas law is in sharp contrast with the chaotic situation that, according to the kinetic theory, exists inside a gas, where countless molecules colliding with one another move in all directions, in zigzag fashion.

If the velocities of all the molecules could be specified at a given instant, and if the interactions among the molecules were known, then, *in principle*, the entire future course of the molecules could be calculated using Newton's laws. In practice, however, this is totally impossible because of the huge number of molecules, even in a small volume of gas.

Taking advantage of the very fact that so many molecules are involved, a solution was found by adopting a statistical approach applying the mathematical laws of probability. This approach is the basis of "statistical mechanics". It cannot tell us what happens to individual molecules, but it can derive very useful conclusions about their collective behavior.

When the statistical behavior of the gas molecules inside a cylinder is analyzed mathematically, and the result is compared with the ideal gas law, a very important fact emerges. The external property of temperature is related internally to the random motion of the molecules: the absolute temperature of a gas is proportional to the average kinetic energy of its molecules. We talk of average kinetic energy because the molecules in their random motion have speeds that vary over some range of values.

Heat then is the kinetic energy of particles in random motion, and temperature is a measure of their average kinetic energy. The greater the molecular agitation, the higher is the temperature we measure.

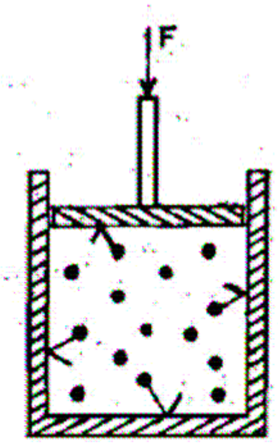

Figure 6.1 - Gas under pressure in a cylinder with a movable piston

If a force bears down on the piston, the piston applies to the gas a "pressure" measured, for instance, in pounds per square inch. The piston will settle at some position where the pressure exerted by the piston on the gas is balanced by the pressure exerted by the gas against the piston. Pressure is also exerted against all parts of the cylinder with which the gas is in contact.

Empirically, Boyle discovered that, provided the mass and the temperature of the gas are kept constant, the product of the pressure times the volume of the gas in the cylinder remains constant, as different forces are applied to the piston. This means that, if we double the pressure, for instance, the volume of the gas is reduced to half its previous value; if we halve the pressure, the volume doubles.

The kinetic theory provides an easy explanation of "Boyle's law". The pressure by the gas against the walls of the cylinder results from countless molecules in random motion continually colliding with, and bouncing away from, either the walls or other molecules. If the volume of the gas is reduced to, say, half, the molecules are packed more closely so that there are twice as many of them per cubic inch as before, anywhere in the cylinder. They will strike the walls of the cylinder, on the average, with the same impact as before (provided the temperature is kept constant), but twice as many collisions will occur, thus doubling the pressure.

Brownian Motion

As late as the first decade of the 20th century, the atomic nature of matter was not unanimously accepted. Among the dissidents were distinguished scientists who argued that it was not necessary to assume the existence of atoms.

Direct conclusive evidence was finally provided around 1908 when the French physicist Jean Perrin, using an ultramicroscope, confirmed experimentally Albert Einstein's theory for the so-called "Brownian motion", named after the Scottish botanist Robert Brown. In 1827, Brown observed under a microscope a rapid jiggling motion of pollen grains suspended in water. Later, he helped to demonstrate that the jiggling motion was not due to any living organisms.

Using statistical mechanics, Einstein was able to develop a quantitative theory, which attributed the jiggling of microscopic particles to their being continuously bombarded by the much smaller molecules of the medium in which the particles were suspended.

To convey some idea of how small and how numerous atoms are, Lord Kelvin used the following example. "Suppose that you could mark all the molecules in a glass of water; then pour the contents of the glass into the ocean and stir the latter thoroughly so as to distribute the marked molecules uniformly throughout the seven seas; if then you took a glass of water anywhere out of the ocean, you would find in it about a hundred of your marked molecules." [1]

Chapter 7

WAVES

The concept of waves was as dominant in physics in the 1800's and 1900's as the concept of particles was in the two preceding centuries. A familiar example of a wave is provided by a long row of upright dominos. If the first domino is tipped over, it pushes the domino in front of it, which in turn pushes the next one, and so on, until all the dominos have been knocked down. We have a disturbance that can propagate quite far, even though each domino moves very little.

We can generate a different kind of waves by holding one end of a long horizontal rope, whose other end is fastened to a support. By flipping our hand up and down, we can cause a series of ripples to snake their way toward the other end of the rope. The way we flip our hand determines the shape of the ripples or waves we generate. A wave shape of particular interest is the so-called *sinusoidal* wave depicted in **Figure 7.1**.

Note that, from instant to instant, it is different points on the rope that are at maximum up, or down, displacement. What we see is not a stationary pattern, but one that snakes its way along the rope at some speed. These waves are said to be "transverse", because each segment of the rope is displaced in a direction perpendicular to the direction of propagation.

Depending on how we move our hand, we may or may not generate ripples that lie all on a single plane, vertical or otherwise. If we do, we say that we are generating "polarized" transverse waves.

Ripples on a water surface

More complex patterns occur when waves propagate on a plane, or in all directions in space. If a small pebble, for instance, is dropped on a very calm pond, it creates a disturbance that propagates radially in all directions on the water surface, generating a circular pattern of spreading ripples.

If, instead of a small pebble, we drop a long, thin stick, we get a different pattern called a "line wave". The disturbance now propagates in the direction perpendicular to the stick. Instead of spreading concentric circles, we see spreading parallel lines resembling ocean breakers rolling onto a beach.

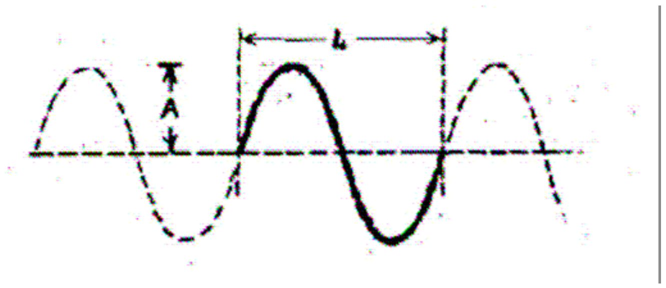

Figure 7.1 - A sinusoidal wave

The curve above displays what happens, at one particular instant, at all points along the line of wave propagation, the rope in our example. (The same curve can be seen also as representing what happens, at one particular point, at subsequent instants of time.)

We see a series of waves, each consisting of a crest above, and a trough below, the base line. The maximum displacement A, above or below the line, is called the amplitude of the wave.

The distance L between any two corresponding points of two consecutive waves is called the wavelength. It defines a cycle that is repeated again and again. Points that are one wavelength apart have identical displacements. Points that are half a wavelength apart have equal but opposite displacements.

The number of waves generated per second is called the frequency. (This is the number of up-and-down ripples we generate in one second by flipping our hand.) Depending on its magnitude, frequency is expressed in Hertz, cycles per second, kiloHertz (thousands of Hertz), megaHertz (millions of Hertz), etc.

The speed of propagation is the distance traveled by the waves in one second. It is equal to the length of a single wave (the wavelength) times the number of waves generated per second (the frequency)

.

Interference and Diffraction

Two phenomena that are most characteristic of the behavior of waves are "interference" and "diffraction".

Since waves have ups and downs, two waves acting on the same point(s) can cancel or reinforce one another, a phenomenon known as "interference". Consider two identical sinusoidal waves (remember Figure 7.1). If they exactly overlap one another, or are displaced from one another by a whole number of wavelengths, their interference will create a combined wave whose crests are twice as high, and whose troughs are twice as low. On the other hand, if two identical sinusoidal waves are displaced by half a wavelength, or an odd number of half wavelengths, they will completely cancel one another.

Diffraction, on the other hand, is the spreading of waves around obstacles, as illustrated in **Figure 7.2**. An interesting combination of diffraction and interference patterns that will be of importance later is shown in **Figure 7.3**.

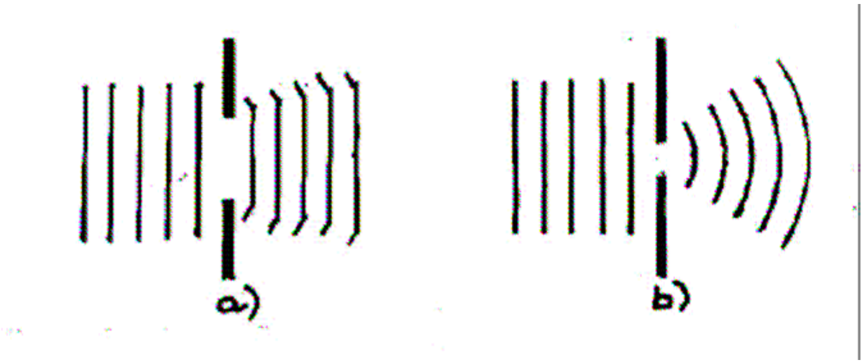

Figure 7.2 - Diffraction of line waves

a) On a water surface, a train of parallel line waves approaches from the left a barrier with an opening. If the size of the opening is sufficiently larger than the wavelength of the approaching waves, the waves generated on the other side of the barrier are also line waves of the same wavelength but with a slight rounding at the edges.

b) As the size of the opening becomes smaller and smaller, relative to the wavelength of the oncoming waves, the rounding becomes more and more pronounced, until the waves generated on the other side of the barrier become semicircular waves with the same frequency and wavelength as the oncoming line waves.

Standing Waves

In addition to *traveling* waves such as those considered so far, there are also *standing* waves, which oscillate in place without traveling. Consider, for instance, a taut guitar string whose end points are fixed. If the string is plucked at its midpoint, it will start vibrating. The result is a single standing half-wave. The amplitude of the up-and-down oscillations is largest at the center; it decreases on either side as we go toward the two fixed end points.

A peak gradually turns into a valley, which then gradually turns into a peak, and so on, as the string vibrates. By plucking the string at other appropriate points, we can generate standing waves displaying two, three or more half-waves.

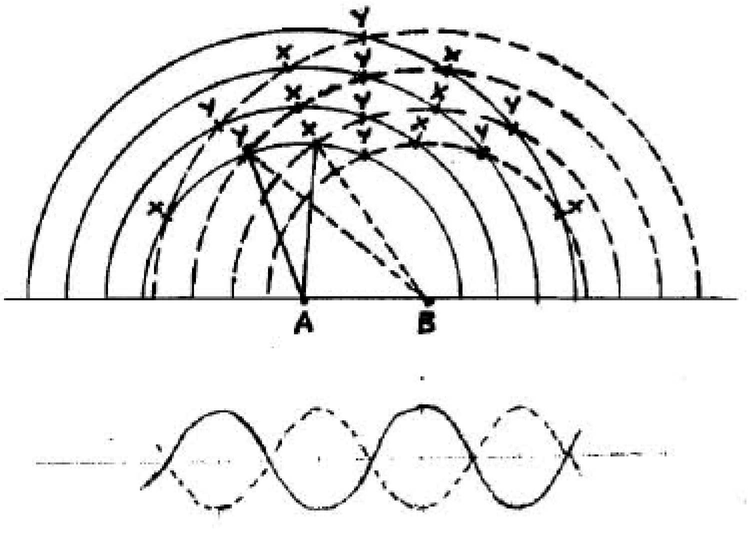

Figure 7.3. Diffraction and interference of identical waves

On a water surface, two suitably small openings A and B create on the other side of a barrier two identical sets of diffraction waves, which interfere with one another.

There are points of zero displacement, such as those marked X, where the waves from A and B always cancel one another. (In the lower part of the figure, we see two waves that have equal but opposite displacements and thus cancel one another.) There are points of maximum displacement, such as those marked Y, where the two waves always reinforce one another to double their values (whether up or down).

Chapter 8

LIGHT

More information by far reaches the human brain through the eyes than through any other sense organ. The carrier of this information is what we call light, a "something" that can be detected by its effects on the retina of the eye, or on a photographic film. But, what is this "something", and how does it behave? The first question has led to mystery after mystery, but centuries of observations have provided a wealth of answers to the second question.

Some General Properties of Light

Whatever it is, light seems to propagate in straight lines: sharp objects have sharp shadows. Strangely, light - the messenger of information - is itself invisible. When we do "see" a beam of light in the darkness, it is only because suspended particles reflect the light. Light travels through air as well as the vacuum of outer space. It travels through some substances, such as glass or water, but not through others, such as metals.

When a beam of light strikes a polished surface, such as a mirror, it is "reflected" like a billiard ball bouncing off a pool table's side.

When a light beam goes from, say, air to water, some of the light is reflected, and some passes into the water, where it is "refracted" (bent in direction). Because of refraction, a straight stick appears bent when partially immersed in water.

Speed of light

Because its speed is so unimaginably high, light appears to travel instantly from source to observer. In the 17th-century, the Danish astronomer Olaf Roemer was the first to estimate the speed of light, by studying one of Jupiter's satellites. This satellite is periodically eclipsed as it moves behind its parent. Roemer noticed that the time between eclipses became shorter as Earth moved closer to Jupiter, and then longer, as the two planets moved farther apart.

He correctly concluded that the discrepancies were due to the time taken by light to cross the distance between the two planets. Over half a year, as Earth moved from a point on its orbit closest to Jupiter to a point diametrically opposite, the total discrepancy had to be the time taken by light to move across the diameter of Earth's orbit. In 1676, Roemer announced that, according to his observations, the speed of light was 140,000 miles per second!

Actually, more accurate modern measurements show that the speed of light is about 186,000 miles per second. Light does not propagate instantly. This fact becomes very important with the immense distances of outer space. Looking into a telescope, an astronomer may observe events that actually occurred millions or even billions of years ago.

Dispersion

As previously mentioned, Newton was the first to observe that the refraction of white light through a glass prism results in its "dispersion" into the colors of the rainbow.

Figure 8.1 shows a beam of white light striking a prism. What emerges is a divergent, multicolored beam. This is due to the fact that the various components of light are refracted differently.

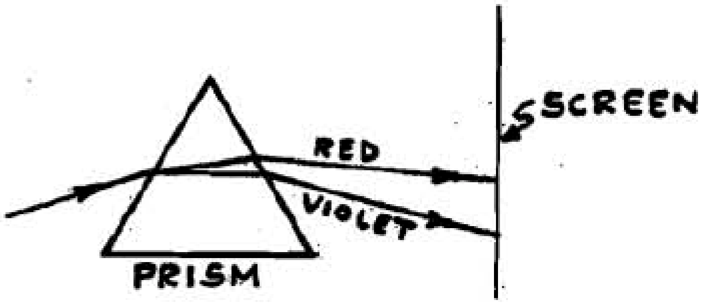

Figure 8.1 - A beam of sunlight through a prism

The composition of white light is more complex than suggested by the prism experiment. The colors of the rainbow from red to violet are merely those components of light that can be detected by the retina of the human eye. Photographic films of various types can "see" additional components that are invisible to the eye: "infrared" components below the red end of the visible spectrum, "ultraviolet" components above the violet end.

Theories of Light: Particles vs. Waves

In Newton's time, there were two theories that attempted to explain the nature of light. One viewed light as a hail of tiny particles; the other, as a wave phenomenon.

In 1690, the Dutch mathematician, astronomer and physicist Christian Huygens presented a theory of light as the propagation of waves. These moved at Roemer's incredibly high speed through "ether", a medium assumed to consist of tiny particles, pervading the entire universe.

In Huygens' wave theory, each particle of the medium did not move much itself, but passed its movement to the next particle in the direction of propagation. (Huygens' waves are said to be "longitudinal" because the displacement of particles occurs in the direction of wave propagation.) In the particle theory, instead, light was seen as a stream of tiny particles, each of which actually moved across great distances.

Newton supported the particle theory and, in the century after his death, his great authority was invoked to support that theory. It was not until the 19th-century that the competing wave theory became firmly established. In 1801, the English physicist and physician Thomas Young performed an experiment that was crucial in proving the wave nature of light. See **Figure 8.2**.

Starting about 1816, Augustin-Jean Fresnel, working independently in France, developed his own wave theory of light, which predicted such effects as diffraction and interference. The key point made by Young and Fresnel in support of a wave theory was that adding one beam of light to another can produce darkness. It is difficult to imagine how particles can cancel one another: by adding particles to particles, one should get more particles. Waves, instead, can cancel as well as reinforce one another.

Fresnel proposed that light waves are "transverse" waves, i.e., with displacements perpendicular to the direction of propagation, as in the waves generated in Chapter 7 by flipping a hand holding one end of a rope. The hypothesis of transverse waves made it possible to account for known phenomena of light polarization. The essence of these phenomena is that, when a beam of light passes through crystals of certain types, its behavior can be affected by the orientation of the crystal. Light that can go through, if the crystal has one orientation, is totally blocked, if the crystal is rotated by 90 degrees.

The hypothesis of transverse waves, however, had some appalling implications. A displacement in the direction of propagation (as in Huygens' longitudinal waves) can be supported by a gas or a liquid. Particles, which are being pushed from behind, in turn push neighboring particles in front, in a pattern that repeats along the direction of propagation. A displacement perpendicular to the direction of propagation (as in transverse waves), however, seems possible only in a solid, where particles are linked together.

For light to be able to propagate through the emptiness of outer space at tremendous speed as a transverse wave, it was necessary to postulate a medium, the ether, which was as tenuous as vacuum, pervaded all space, could penetrate almost any substance, and was not only elastic but solid!

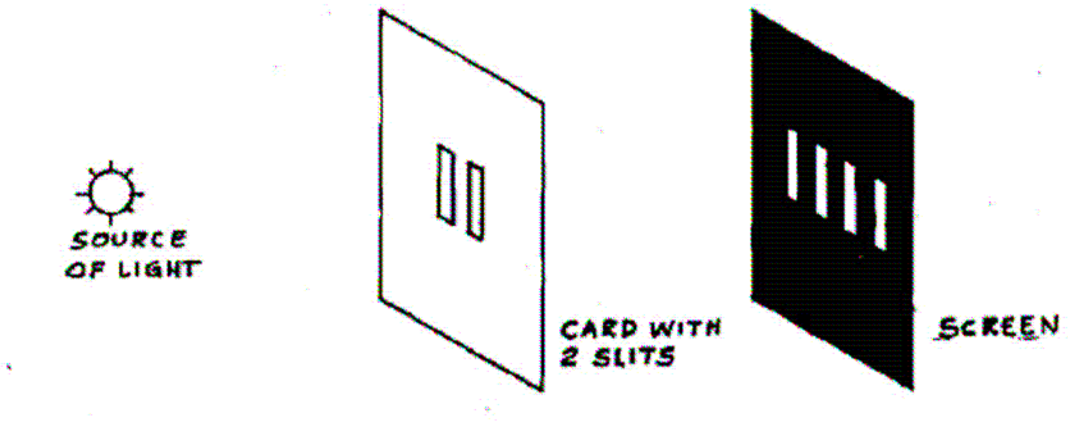

Figure 8.2 - Young's experiment

As shown schematically above, a beam of light is shined in the dark on a barrier, a card with two very narrow slits close together. The light that goes through the slits falls on a screen. If light consisted of tiny particles, we would expect the screen to be totally dark except for two illuminated areas, where the particles that go through the slits strike the screen. What we see, instead, is an alternating pattern of several dark and light fringes (the figure shows only a few, and much sharper than they should be).

The light waves shined on the barrier are diffracted by each of the two slits. On the other side of the barrier, because of diffraction, the two slits act as two identical sources of light waves (remember Figure 7.3). Two sets of waves spread out in a semicircular pattern, interfering with one another. There are points where the waves from the two slits reinforce one another. On the other hand, there are points where the two waves cancel one another. As a result, on the screen we see a series of dark and light fringes.

(From Blanpied Figure 9-24 pg 375)

With a transverse-wave theory, a wide variety of optical phenomena could be explained in a consistent way. At the same time, the new theory raised some very perplexing questions: What was the nature of the "displacement" associated with a light wave? What was it that was displaced? The experiments of Young and Fresnel, however, were found conclusive at the time, and light came to be recognized as a wave phenomenon. This was not to be, however, the last word about the nature of light.

Chapter 9

ELECTROMAGNETISM

In ancient times, it was known that, after being rubbed, amber could attract feathers and other light objects. It was also known that iron was attracted by some mineral from the province of Magnesia in Greece. The magnetic compass, an invaluable aid to navigation, is the oldest practical application of magnetism, but its origins are unknown.

In 1600, Sir William Gilbert (1540-1603), a contemporary of Kepler and Galileo and physician to Queen Elizabeth I, published his book "On the Great Magnet of the Earth", which made him famous throughout Europe. He had spent 17 years experimenting with magnetism and, to a lesser extent, with electricity. He correctly concluded that the Earth acts as a huge magnet, which explains why the needle of a magnetic compass always points toward the North. He also showed that many common materials, when rubbed, behaved like amber. From the Greek word for amber, he called such materials "electrons", from which the word electricity was later derived.

CHARGES AT REST

An amber rod that has been rubbed is said to have become (electrically) "charged", or to have acquired an (electrical) "charge". Many other substances - such as rubber, glass and most crystals - can be charged by rubbing.

If two charged amber rods are brought close together without touching, they repel one another. The same happens with two (charged) glass rods. However, if we bring close together an amber rod and a glass rod, they attract one another. Thus, there are two types of electric charges, which are called positive and negative. Matter is normally neutral because it contains equal amounts of opposite charges that cancel one another.

All matter contains huge numbers of electrically charged particles that are extremely small and close together. In an ordinary piece of matter, the balancing of positive and negative charges is so nearly perfect that it is difficult to observe any electrical force. That is why rubbing was so important in the early observations of electrical effects: it causes a separation of positive and negative charges, altering their precise balance and revealing the presence of a "net" (positive or negative) charge. Like mass, charge cannot be created nor destroyed.

In 1785, a French physicist, Charles Augustin de Coulomb, formulated the law that describes the "electrostatic" force between two stationary *net* charges:

- Its magnitude is directly proportional to the product of the two (net) charges, and inversely proportional to the square of their distance.
- Its direction is along the line connecting the two charges.
- It is a repulsive force (away from the other charge), if the two charges have the same sign; it is an attractive force (toward the other charge), if the two charges have opposite signs.

In spite of the striking similarity between Coulomb's law and Newton's law of gravitation, there are very substantial differences between the electrostatic and gravitational forces:

- The electrostatic force is enormously more powerful than the gravitational force.
- Since charge can be either positive or negative, the electrostatic force can be either attractive or repulsive. On the other hand, only attractive gravitational forces have ever been observed.
- Gravitational forces are felt even at enormous distances throughout the universe because they never cancel out. On the other hand, even powerful electrostatic forces are not felt at all, if they cancel out. Two people standing at arm's length from one another feel no force between them. Yet, if the positive and negative charges contained in their bodies were equally out of balance by just one percent, the repelling force between the two bodies would be large enough to lift a weight equal to that of the entire earth! [1]

FARADAY AND FIELDS

Michael Faraday (British, 1791-1867) was a self-taught scientist with no formal education and only a limited knowledge of mathematics. Yet, he became one of the greatest experimental physicists of all times. He was born near London of a very poor family. After he became an apprentice to a bookbinder at age 14, in the hours after work, he would read some of the books brought in for rebinding.

His great opportunity came at age 21, when a customer gave him tickets for a series of lectures by the renowned chemist Sir Humphrey Davy. After attending the lectures, Faraday expanded the careful notes he had taken, and bound them into a book, which he sent to Davy with a letter asking for a job. Davy advised him not to give up a skilled trade for something in which there was neither money nor opportunity for advancement. A few months later, however, when one of his laboratory assistants had to be dismissed, Davy remembered Faraday and offered him a job. Thus began one of the most illustrious careers in the history of science, with many important contributions first in chemistry and then in physics.

As a reward for his lifetime devotion to science, Queen Victoria granted him the use of a house, and offered him a knighthood. Faraday gratefully accepted the house, but refused the knighthood. He wanted to remain, he said, plain Mr. Faraday to the end.

The Electric Field

Faraday - who relied on mental visualization, rather than mathematical abstraction - introduced one of the most important ideas in physics, the concept of *field*, as an aid to visualize the effects of charges on other charges.

Coulomb's law carries the same action-at-a-distance implication of Newton's law of gravitation, an implication that Faraday, like many others, strongly rejected. He proposed, instead, that a charge placed at some point creates a field of influence that extends throughout space. A small test charge placed at some other point is affected by the "field value" that exists there, and not by some action-at-a-distance from the first charge.

If we have a number of charges at various points, they create a single field that represents the combined effect of all the charges. By definition, the electric field at any point is the force per *positive* unit charge. The force acting on a charge at that point is then the value of the charge times the electric field.

As proposed by Faraday and illustrated by the two examples in **Figure 9.1**, an electric field can be depicted by drawing "field lines", or "lines of force". Where the lines of forces are more concentrated, the forces are stronger. As charges change or move, the whole pattern of lines of forces changes accordingly.

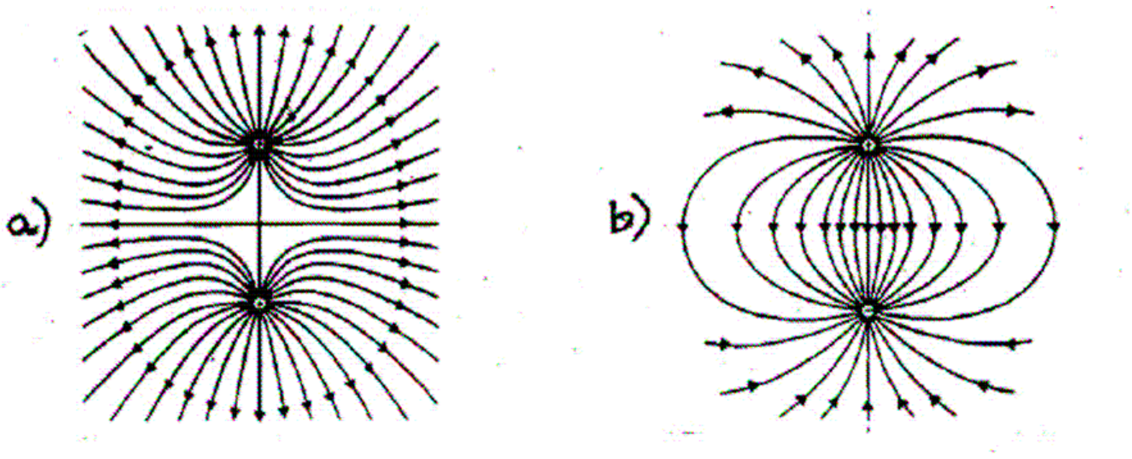

Figure 9.1 - Two examples of electric fields created by two equal charges.

In a), both charges are positive; in b), the upper charge is positive; the lower is negative. A small positive test charge placed at any point in a field will be impelled in the direction of the field line passing through that point. (The test charge is assumed to be small enough not to alter the pre-existing field.) The field varies in value and direction from point to point.

The Magnetic Field

Magnetic forces were first observed in connection with natural magnets. Magnets, like the familiar horseshoe magnet, have two end regions, called the "poles", where the magnetic forces are strongest. The two poles of a magnet are not alike. If a magnet is suspended by a thread, it will align itself with the earth's magnetic field, one pole pointing to the North, the other to the South. Two like poles (both north or both south) always repel one another, whereas opposite poles always attract one another.

The influence exerted by a magnet can also be described in terms of a "field" in the surrounding space. This magnetic field can be depicted by means of field lines, or lines of force, as illustrated in **Figure 9.2**.

To explain electric and magnetic effects, even gravity, Faraday used mental pictures of lines of forces reaching out through empty space and interacting with one another. He viewed atoms not as tiny lumps of solid matter, but as centers of concentration of forces.

MORE ABOUT THE ATOM

It will be easier to discuss the discoveries in electromagnetism that unfolded in the 1700's and 1800's, if we take now a peek at the internal structure of the atom, which was not discovered until the late 1800's and early 1900's.

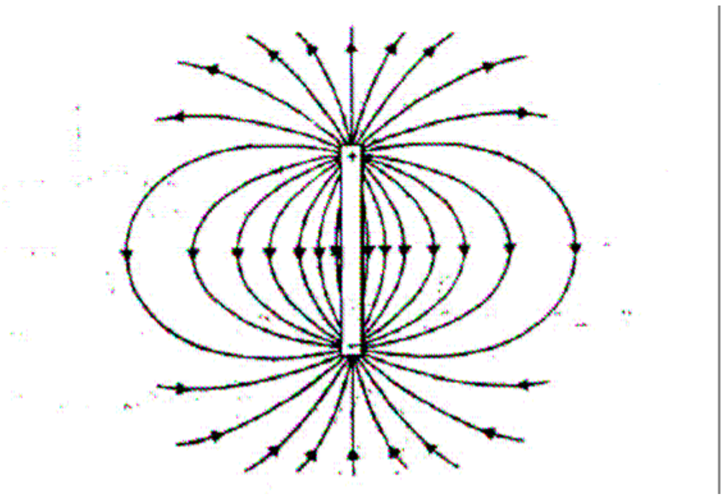

Figure 9.2 - Magnetic field due to a bar magnet

A small test magnet placed at any point in the magnetic field created by the bar magnet above will orient itself in the direction of the field line passing through that point.

In a familiar experiment, iron filings scattered around the bar magnet on a horizontal plane will arrange themselves in the pattern of the field lines depicted above.

In the classical atomic theory of matter, the atom was the ultimate indivisible particle. Actually, it has a complex internal structure. For the time being, it will be helpful to view the atom as if it were a tiny planetary system with a positively charged nucleus around which orbit a number of much lighter particles, called "electrons", all with identical negative charges. The positive charge of the nucleus always equals the combined negative charge of all its orbiting electrons, thus making the atom neutral (zero net charge).

Between the nucleus and an electron, there are both gravitational and electrical forces of attraction. The electrical force, however, is a billion billion billion billion times larger than the gravitational force. So dominant at the cosmic scale, the gravitational force is totally dwarfed by the electrical force at the atomic scale.

"The force that holds the atoms together, and the chemical forces that hold molecules together, are really electrical forces acting in regions where the balance [between positive and negative charges] is not perfect, or where the distances are very small."[2] Within a sample of matter, atoms are held together by electrical bonds. Several types of such bonds can be formed, depending on the nature of the atoms involved. In materials called "insulators", all the electrons are restrained from straying from their parent nuclei.

In materials called "conductors", instead, one or more "free electrons" from each atom are able to move about, whereas the remaining electrons are kept bonded to their parent nuclei. The positive nuclei and their captive electrons are fixed in a 3-dimensional grid.

The "free electrons" are repelled by surrounding negative charges; at the same time, they are attracted by the surrounding positive nuclei. As a result, moving in all directions, they experience many collisions, almost continuously changing direction.

CHARGES IN MOTION

Electric Current

By definition, any motion of net charge constitutes an electric "current". Of particular interest is the flow of a current through a metal wire. The current is defined as the amount of charge that passes through the cross section of the wire in a unit time. A unit of measure for current is the familiar "ampere" or "amp", named after the French physicist Andre-Marie Ampere (1775-1836). One ampere of current corresponds to the flow of about ten billion billion free electrons per second through the cross section of the wire.

An external source of energy is required to maintain a net flow of free electrons; it is called a source of "EMF" or "electromotive force". It performs a function analogous to that of a mechanical pump in a closed loop of water pipes. The operation of the pump can be characterized by the volume of water it can displace per unit time (whose electrical counterpart is the electric current), and by the pressure differential the pump can establish between its inlet and outlet. The electrical counterpart of the latter is called "voltage", which is measured in "volts", named after the Italian physicist Alessandro Volta (1745-1820).

In the presence of a source of EMF, an electric field is established along the wire affecting the motion of the free electrons. A consistent drift is now superimposed on their random motion. In spite of the many collisions and resulting changes in direction, more electrons cross in one direction than in the other, and a net current flows through the wire. Because of all the many collisions, part of the energy supplied by the source of EMF is dissipated in the form of heat.

The most common sources of EMF are batteries and generators. The battery was invented in 1800 by Alessandro Volta. It remained the main source of current until the first practical electric generator was built in the late 1860's.

Magnetic Effects of Currents

After the invention of the battery, the availability of a source of continuous current led to important discoveries. A major turning point occurred in 1820, when a Danish physicist, Hans Christian Oersted, announced that electric currents have *magnetic* effects. He made this discovery during a class demonstration to his students when, by accident, he placed a wire that carried current near a magnetic compass, and was surprised to see the needle swing to a direction perpendicular to the wire.

Here was an unprecedented case of magnetism without a magnet! What makes the discovery all the more intriguing is that a stationary charge, however large, has no effect on a magnetic needle. A charge must be *moving* in order to create a magnetic field. The larger the charge, and the faster its motion, the stronger the magnetic field it creates.

After Oersted's discovery, the magnetic effects of an electric current were studied by Ampere and others. (By the age of 12, Ampere had mastered all the mathematics then known.)

As an example, **Figure 9.3** shows the magnetic field that is produced near the middle of a long straight wire carrying current. At any point, it is proportional to the current, and inversely proportional to the perpendicular distance to the wire.

Induced Currents

When it became known that an electric current produces magnetism, it seemed natural to expect that magnetism, in turn, should produce a current somehow. It would have been of enormous practical importance to find a way of producing a current other than by the use of a battery.

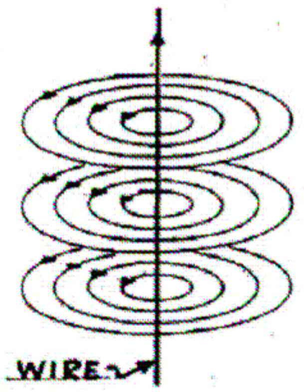

Figure 9.3 - Magnetic field lines due to a current through a straight wire

A magnetic field coils itself around a wire-carrying current, as shown above. For a circuit of a different shape, for instance, a circular loop, the magnetic lines of force will coil around the wire in a doughnut-like pattern.

If the vertical wire above goes through a perpendicular plane with iron filings scattered on it, they will arrange themselves along the lines of force.

Faraday worked on and off for 10 years trying to prove that magnetism could induce a current. He finally succeeded in 1831.The underlying principle of Faraday's discovery can be stated qualitatively as follows: *a changing magnetic field produces a changing electric field.* The electric field can be produced not only along a conducting wire, where it can cause a flow of current, but also in empty space, even in the absence of wires and charges.

Before, we saw that a charge will induce a magnetic field, *but only if it is moving*. Now, we see that a magnetic field will induce a current, *but only if it is changing*.

When Faraday first announced his discovery, he was asked "What is the use of it?" Faraday replied: "What is the use of a new-born baby?" That baby grew to create our modern electric power industry with its dams, power plants, transformers, miles and miles of high-tension wires etc.

MAXWELL AND ELECTROMAGNETIC RADIATION

James Clerk Maxwell (1831-1879) ranks alongside Newton and Einstein for the fundamental nature of his contributions. He is best known for his brilliant mathematical theory of electromagnetic radiation.

The circumstances of Maxwell's life were quite different from those of Faraday's. He was born in Edinburgh, Scotland, the son of a lawyer. When he was only 8, his mother died at 48 of abdominal cancer, the very disease that would claim his own life at exactly the same age. In 1854, he obtained a mathematics degree from Trinity College in Cambridge.

In 1860, at the age of 29, he was appointed to the professorship of natural philosophy (as physics was still called) at King's College in London. During the next five years, he developed his theory of electromagnetic radiation. In 1865, he retired to his family estate, devoting most of his energies to writing his "Treatise on Electricity and Magnetism", which was published in 1873. In the preface, he stated that his major task was to put Faraday's intuitive ideas about fields into mathematical form.

When Maxwell died in 1879, unlike Newton, he received no public honors, and was buried quietly in a small churchyard. Although his theory of electromagnetic radiation "remains for all time one of the greatest triumphs of human intellectual endeavor", he is still unknown to many people.

Maxwell's Equations

Maxwell wrote a number of equations to state in mathematical form what was known at the time about electricity and magnetism. In overly simplified fashion, we can summarize it as follows:

- Electric charges, which may be positive or negative, create an electric field.
- A current (a charge in motion) produces a magnetic field.
- A changing magnetic field induces a changing electric field.

Could a changing electric field, in turn, induce a changing magnetic field? Maxwell theorized that it could, and added accordingly one more term to his equations. Maxwell's modified equations predicted that a changing current (a charge in accelerated motion) initiates the propagation of an electromagnetic wave. The process involved can be summarized, *approximately*, as follows:

- At a transmitting source, a changing current induces a changing magnetic field.
- A changing *magnetic* field induces a changing *electric* field.
- Conversely, a changing *electric* field induces a changing *magnetic* field, which then induces a changing *electric* field, and so on.

The electric and magnetic fields work their way through space - "pushing" each other along. As a result, an electromagnetic wave propagates through space, and is capable of inducing a current at some distant receiver, concluding a chain of events initiated by the changing current at the transmitting source.

Electromagnetic radiation is produced whenever a charge is in accelerated motion (i.e., its speed and/or direction changes). A charge that is oscillating back and forth, for instance, or moving along some closed loop, radiates electromagnetic energy.

Let us assume that a changing current has set in motion the propagation of an electromagnetic wave. If the current were to stop flowing, the waves initially created would continue to spread. The electromagnetic field at some point in space depends only on the field in the *immediate* neighborhood

If the current at the transmitter alternates continuously in sinusoidal fashion, it generates a continuous train of electric and magnetic waves. The waves represent changing values of a magnetic and an electric field propagating in some direction. At any point, the electric field and the magnetic field are perpendicular to one another and to the direction of propagation.

From his equations, Maxwell was able to calculate a propagation speed for electromagnetic waves. He found that it agreed very closely with the speed of light, which led him to conclude that light itself is an electromagnetic phenomenon.

Electromagnetic radiation appears in a wide variety of forms, which are all electromagnetic waves of different frequencies, all traveling in free space at the speed of light. They include, in increasing order of frequency, radio- and TV-broadcast waves, microwaves, radar waves, infrared light, visible light, ultraviolet light, X-rays, and gamma rays.

What we call (visible) light consists of electromagnetic waves whose frequencies range between 500 million million cycles per second for red light, and 1000 million million cycles per second for violet light. The corresponding wavelengths are 0.00008 cm. for red light, and 0.00004 cm. for violet light. Note that red light and violet light, for instance, differ only in frequency (or wavelength): color is something that happens in our minds.

As the frequencies of electromagnetic waves go higher and higher, to one million million million million cycles per second and more, the associated wave lengths become shorter and shorter[4], down to one thousand-million-millionth of a centimeter and less!

To understand how such very high frequencies (and very short wavelengths) can be generated, we have to look to the very rapid motions of

---

[4] You may remember that the speed of wave propagation is equal to wavelength x frequency. Since all electromagnetic waves travel in free space at the same speed, the lower their wavelength, the higher their frequency, and vice versa.

charges within atoms, or even within their nuclei.

Experimental Verification of Maxwell's Theory

It was not until about 1887, eight years after Maxwell's death, that the German physicist Heinrich Hertz announced that he had been able to produce electromagnetic waves of the type predicted by Maxwell. Between 1885 and 1889, he generated electromagnetic waves in the laboratory, and studied their behavior. He confirmed experimentally that their speed is equal to that of light, and that they behave just as light does, except that they are not visible. For instance, they can be reflected as well as refracted. Light was indeed a special case of electromagnetic radiation.

Electromagnetic waves, which can travel great distances, can be used to transmit information (send signals), if they are modified in some fashion. One way is to transmit shorter and longer bursts in some telegraphy code of dots and dashes.

In 1901, the Italian inventor Guglielmo Marconi succeeded in receiving signals transmitted by means of electromagnetic waves across the Atlantic Ocean. Signals for the letter s in Morse telegraphy code (three dots) travelled a distance of 2,000 miles from England to Newfoundland, in Canada.

This achievement created an enormous international sensation, and launched the development of radio communications, radio and TV broadcasting, radar etc., all born from Maxwell's equations.

Chapter 10

A PUZZLING INCONSISTENCY

By the latter part of the 19th century, physics was able to explain in detail a wide range of natural phenomena, using only a remarkably small number of principles. Some believed that physics had reached its peak, that Newtonian mechanics and Maxwell's theory of electromagnetism represented the completion of theoretical physics. It seemed that nature held no more secrets of importance!

Actually, between 1900 and the late 1920's, the very foundations of physics would change so radically that many historians describe this period as a scientific revolution comparable to the one that took place during the 16th and 17th centuries. Two fundamentally new theories emerged during this period: relativity and quantum theory, the two pillars on which contemporary physics rests.

Newton's and Maxwell's equations described two magnificent theories, each brilliantly successful in its own field. It was recognized, however, that there was a basic inconsistency between the two. It was natural to suspect first Maxwell's equations, which were only decades old, rather than Newton's equations, which had ruled unchallenged for some 200 years.

The solution came from Albert Einstein's drastic reevaluation of the concepts of space and time at the very roots of Newtonian mechanics. It is these concepts we need to review now, before we can see the nature of the inconsistency between Newton's and Maxwell's laws.

Frames of Reference

To a passenger sitting in a plane flying at 600 miles per hour, a stewardess appears to move around quite normally, just as if the plane had never taken off. To somebody on the ground, however, she appears to be rushing at high speed.

What is observed in nature is always relative to the particular "frame of reference", from which the observing is done (such as from the ground or from a plane). We might think of a reference frame as a "platform" carrying an observer equipped with some reference from which to measure distances, a clock to measure time, and whatever else might be needed to study the laws of physics.

Space and Time in Newton's Mechanics

Newtonian space is conceived as a vacuum, or empty space, within which particles move and interact, like fish swimming inside a water tank. Space is considered to be infinite and uniform. It is a passive, unchangeable stage that neither affects, nor is affected by, the presence and motion of matter.

In principle, regardless of where they are located and whether they are moving or not, any two observers can always agree on their measurements of distance. Moving or not, a yardstick represents the same length anywhere in the universe.

Similarly, time is thought to be absolute and universal. We can talk of a particular "moment in time" that is the same everywhere in the universe. Events that occur anywhere at the same "moment in time" are simultaneous and, in principle, can be so recognized by everybody. At least ideally, any two reference frames can have clocks that go at exactly the same rate. If some physical process takes, say, one hour, as determined in one reference frame, it will take precisely one hour with respect to any other frame, moving or not.

Many of the Newtonian notions just reviewed are very much part of our intuitive way of thinking about reality. That time flows equally everywhere for all observers, regardless of their state of motion, is so deeply ingrained in our minds that we cannot even imagine an alternative.

Inertial Frames of Reference

Newton's first law of motion, the law of inertia, states that, in the absence of any resistance, an object moving at some speed in some direction will continue to move at the same speed and in the same direction, without the intervention of any force. We say that the object continues to move on its own "by inertia". Only if some force starts acting on the object, will the object respond by changing its speed and/or direction.

The law of inertia raises the question: What frame of reference do we have in mind, when we talk of a state of rest, or of uniform motion? If somewhere in the universe one could find an object absolutely at rest, a "fixed center of the universe", Newton's laws would be valid with respect to a "platform" attached to that object. From it, an observer could measure absolute positions and absolute speeds anywhere.

We know, however, that our Earth is rotating about its axis and revolving around the Sun. The Sun, in turn, is revolving about the center of the Milky Way; at the same time, our galaxy is moving as a whole. With all this motion in the universe, it seems hopeless to determine whether anything at all is at rest. In practice, we can deal only with relative positions and relative motions. Absolute positions and absolute motions are unobservable.

Newton's laws are valid also on any platform that, without rotating, moves uniformly (with constant speed and direction) with respect to our hypothetical "fixed center of the universe". Any such platform, or frame of reference, is said to be "inertial". Any inertial frame can claim to be itself the "fixed center of the universe", because Newton's laws are valid with respect to it.

Note that the Earth is not an inertial frame of reference, because it rotates about its axis and revolves around the Sun. Thus, strictly speaking, Newton's laws are not valid on such a reference frame. Solutions to most engineering problems, however, can be obtained with satisfactory accuracy, if we view an Earth-bound reference frame as if it were inertial.

The Special Principle of Relativity

In the Principia, Newton stated what is now called the "special principle of relativity" or the "Galilean relativity principle":

> *"The motions of objects contained in a given space [such as a room] are the same relative to one another, whether that space is at rest, or moving uniformly in a straight line without any circular motion."*

It is called the "special" principle because it is valid only if the "room" is at rest, or moving uniformly in a straight line, without rotating.

Let us imagine, for instance, two trains on separate tracks fixed to an inertial "platform": their tracks go on for miles and miles, perfectly straight. On each train, a physicist sets up a laboratory in a car without windows, to carry out various experiments. Each train may be at rest or moving at constant speed. When the train moves, the ride is perfectly smooth.

The special principle of relativity tells us that, inside the two windowless laboratories, all experiments performed will appear exactly the same to our two physicists. For instance, upon dropping a ball, each will observe that it falls straight down. If he throws the ball straight up, shortly after, the ball will start falling straight down toward his feet.

Whether moving or not, everything looks perfectly normal to each physicist. Since he cannot look out, *he has no way of telling whether he is moving or not*, because the same laws of mechanics hold in both trains, whether at rest or in uniform motion.

A stowaway inside a ship with a very quiet engine, on a very smooth ocean, would not be able to tell whether the ship is moving or not. Looking up, he might see clouds going by, but could still wonder whether the ship or the clouds were moving.

It is only uniform velocity that cannot be detected without looking out. Even inside an enclosed space, without looking out, we can detect an acceleration, i.e., whether we are changing speed and/or direction, because we feel pulled or pushed in some direction. If the change is sudden enough, we might even be thrown against a wall.

Views from Two Platforms

Let us compare now observations made by two people on separate inertial platforms. In Newtonian mechanics, measurements of time, distance, speed and acceleration made on two inertial frames are related in a way that appears to us intuitively obvious.

Suppose you are standing by a long straight train track on some inertial platform. There are no trees, buildings or other landmarks that can be used as reference points. You see me standing on a very long flat car (my inertial platform) moving East along the track at 20 miles per hour with respect to you. From my viewpoint, however, I am standing still; you and the track are moving West at 20 m/h.

At some point along the track, you have planted a pole from which you measure distances along the track. I am measuring distances from a pole at the rear of my train.

We both have equally reliable clocks and are always in agreement as to the *time* t elapsed since our reference poles passed by one another. We are both observing some jogger on my flatcar jogging toward the front.

Our measurements of *distances* to the jogger will differ by a changing amount, namely, the changing distance between our two reference poles (which is equal to the constant speed of the train multiplied by the time t).

Our measurements of the jogger's *speed* differ by a constant amount, the speed of the train, 20 m/h. What is 5 miles per hour to me is 25 (20+5) m/h to you. (If the jogger were moving in the opposite direction, it would be 20-5=15 m/h.) This "obvious" way of *combining speeds by addition (or subtraction)* is a key point to keep in mind.

When it comes to *acceleration*, however, we are always in agreement. Whether the jogger, in one second, accelerates from 5 to 6 m/h, or from 25 to 26 m/h, it is the same acceleration, 1 m/h in one second.

Whether on your inertial platform or mine, Newton would have written his mathematical laws in exactly the same form. His laws were invariant to (not affected by) uniform motion. Maxwell's laws of electromagnetism, instead, were not invariant from one inertial platform to another in the way Newton's laws of mechanics were. This is the inconsistency that puzzled physicists in the latter part of the 19th century.

Chapter 11

THE ELUSIVE ETHER

"*One thing we are sure of, and that is the reality and substantiality of the luminiferous ether .... You may regard the existence of the luminiferous ether as a reality of science.*"

Lord Kelvin (1824-1907)

Maxwell's equations predicted a single definite value for the speed at which electromagnetic waves, and light in particular, propagate in a vacuum: c = 186,000 miles per second (c is the letter traditionally used to represent the speed of light).

By Newton's mechanics, the speed of light ought to vary with the speed of the "platform" on which it is measured. Let's say that a beam of light is traveling at 186,000 miles/second with respect to some platform A. If some platform B is traveling in the direction of the beam of light with a speed of 100,000 miles/second with respect to platform A, an observer on B should measure the speed of the beam of light to be 86,000 miles/second. If he knows that platform A has measured 186,000 miles/sec., he can conclude that his platform is moving at 100,000 miles/sec. with respect to platform A.

Using the speed of light derived from Maxwell's equations, one ought to be able to do what was not possible using only Newton's laws of mechanics, namely, to detect uniform motion.

The Ether

The speed of light is 186,000 miles per second. But with respect to what? Earlier, we defined a wave as a disturbance of some kind that propagates through a medium (like the ripples that propagate on the surface of a pond). If light was a wave, it seemed only natural to assume that it too had to have a medium in which to "wave." A medium had to be invented so that physicists could visualize the propagation of light. Then, the speed predicted by Maxwell's equations might be the speed of light relative to this medium, which, as previously mentioned, was called the "ether". This medium was supposed to be some invisible substance filling all space, a form of matter, probably composed of particles.

In the latter part of the 19th century, belief in the existence of the ether became firmly entrenched in the minds of scientists. In Maxwell's words, "there can be no doubt that the interplanetary and interstellar spaces are not empty, but are occupied by a material substance or body, which is certainly the largest and probably the most uniform body, of which we have knowledge."

And yet, as mentioned in Chapter 8, the ether was a most peculiar substance on which to place such total faith! Since the motions of the planets had been observed over millennia without noticing any change, the ether must offer at most an infinitesimal resistance to the motion of the planets. Most likely, the ether could freely penetrate ordinary matter, as "the water of the sea passes through the meshes of a net when it is towed along by a boat", in Maxwell's words. The ether had to be very low in density, and yet very hard, since both of these properties result in a high speed for waves. The ether made it possible to visualize the propagation of light, but was itself most difficult to visualize, because of the contradictory properties it had to possess.

If the ether really existed, the Earth should be moving through it at some definite speed. To an observer on Earth, the ether should appear then to be moving toward the Earth. If we are riding, for instance, in an open car and there is no wind, it feels to us, however, as if a wind were blowing against the car. Similarly, there should be an ether "wind" or "stream" blowing against the Earth.

Physicists felt that the ether "wind" ought to be strong enough to have some observable effect. A number of very ingenious experiments were devised to detect the motion of the Earth through the ether. The first and most famous of these experiments was carried out by an American physicist, Albert Michelson (1852-1931), with the aid of an American chemist, Edward Morley (1838-1923). The results of their experiment were announced in 1887.

The Michelson-Morley Experiment

Born in Germany, Michelson came to the United States with his parents when he was only two years old. In 1873, he graduated from the Naval Academy in Annapolis. Six years later, he took a leave for a few years of study in Germany and France, where he became intrigued by the problem of detecting the ether. After returning to the United States, he resigned his Navy commission to become in 1883 the first physics professor at the newly founded Case Institute in Cleveland, Ohio. In 1907, he became the first American to receive the Nobel Prize for Physics.

[The Nobel Prize is named after Alfred Bernard Nobel (1833-1896), a Swedish chemist, engineer and industrialist, who accumulated an immense fortune from his invention of dynamite and other explosives. Being somewhat of a recluse, he never married; when he died, he left the bulk of his estate to establish what have become the most prestigious international awards. These are given annually to those who "have conferred the greatest benefit to mankind" in the fields of physics, chemistry, medicine, literature, economics, and the promotion of peace.]

To explain his experiment to his children, Michelson compared it to a race between two swimmers. We will talk, instead, of two identical race boats, both set to move at exactly the same speed with respect to the water. Starting from one bank of a river, one boat goes straight across to the other bank, and back. The other boat goes the same distance (as measured from the ground), but in the direction of the river, once upstream and once downstream.

Although the two boats travel at the same speed with respect to the water, they travel at different speeds with respect to the land because of the current. When the boat that travels along the river is going downstream, the speed of the current must be added to that of the boat. When the same boat is going upstream, the speed of the current must be subtracted from that of the boat. The boat that goes across, instead, is slowed down in both directions, because it must always be aimed a little upstream in order to prevent the current from pulling it off course.

Note that, we have combined speeds by addition or subtraction in Newtonian fashion. It can be shown mathematically that the boat that goes across will win every time, because it actually travels a shorter distance with respect to the water. Only if the river were still, would the two boats finish at the same time.

The basic idea of the Michelson-Morley experiment was to set up a race between two beams of light. The whole planet Earth corresponds to the river banks of our boat race, and the ether "stream" corresponds to the flowing river. Instead of the two race boats, we have two beams of light making round trips of equal lengths, but in perpendicular directions, through the ether, which is streaming by the Earth. The beams are then rejoined.

If there is indeed any relative motion between the ether and the Earth (the river and the banks), by Newton's mechanics, one beam should "win the race". This means that the two emerging beams would be out of step to some extent and would interfere with one another. Very ingeniously, the experiment took advantage of the interference of waves to reveal extremely small time differences by means of concentric dark and light interference rings.

To Michelson's great surprise, however, the race had no winner! The observations were continued over a long period of time - with the Earth going in one direction and, six months later, in the opposite direction. The experiment has since been repeated many times in various ways. No motion of the Earth with respect to the ether, however, has ever been observed.

Later, the hypothesis of the ether - as a physical medium in which electromagnetic waves "wave" like ripples on the surface of a pond - was deemed unnecessary, and was abandoned.

The "Conspiracy" of Nature

Various attempts were made to explain the unexpected failure of the Michelson-Morley experiment. Many other experiments were designed to detect an ether stream, but they all failed. It appeared that nature was in a conspiracy to defeat all such attempts: it always seemed to introduce some new phenomenon to undo whatever ought to have made possible the detection of the ether stream.

The French mathematician and theoretical astronomer Henri Poincare' realized that a total conspiracy is itself a law of nature. In the late 1800's, he restated Newton's special principle of relativity more broadly as follows:

The laws of physical phenomena must be the same for a fixed observer A as for an observer B who has a uniform straight motion without rotation relative to A. We do not have, nor can we possibly have, any way of determining whether or not we are carried along in such a uniform straight motion.

There did not seem to be any way of detecting an absolute motion, in particular, that of the Earth with respect to the ether. And yet, knowledge of the speed of light should have made it possible!

Chapter 12

SPECIAL RELATIVITY

Albert Einstein

Eighteen years after the results of the Michelson-Morley experiment were announced, the solution to the puzzle they created came in 1905 from an obscure examiner of the Swiss Patent Office, Albert Einstein, with his "Special Theory of Relativity", which was followed in 1916 by his broader "General Theory of Relativity".

In 1905, at the age of 26, Einstein published four papers, which changed forever our view of the Universe:

- The first paper provided a theoretical explanation of Brownian motion. As previously discussed, it led to wide acceptance of the atomic theory of matter.
- The second paper, in which Einstein proposed a new theory of light, represents a major contribution to quantum theory at its very beginnings. Thus, we find Einstein playing a leading role in both key theories of 20th-century physics.
- The third paper introduced Einstein's special theory of relativity.
- The fourth paper was a follow-up to the third. It established the equivalence of mass and energy in physics' most famous equation: $E = m \times c^2$ (Energy = mass x the square of the speed of light).

There had not been a comparable outburst of scientific creativity since the plague years of 1665-66, when, within a few months, the 23-year old Newton discovered the composite nature of light, analyzed the action of gravity, and invented calculus.

In the 1600's, the shock created by the "Copernican Revolution" had been enormous. Mankind had to adjust to the idea that we live on a planet rushing through space at more than 66,000 miles per hour. In retrospect, however, we must consider this shock rather mild, compared to the violence that Einstein's theory of relativity did to the deeply entrenched notions of time and space. More blows to a common-sense view of reality were to come from the quantum theory, more than even Einstein could accept!

One cannot say about Einstein the usual things that are told about great geniuses - for instance, that the brilliance of his mind was recognized early in his life, or that, in his college days, like Newton, he caught the eye of a professor willing to recommend him for his own job as a man of "unparalleled genius". Like Newton, he was not a precocious child. In fact, he was slow to learn to speak, and his parents feared he might be retarded. In high school, he was told that he would never amount to anything. After he graduated from college, he had difficulty finding a permanent position for lack of recommendations - he later recalled that he had not been in the good graces of any of his former professors (one of them called him a "lazy dog").

He was born in Ulm, Germany, in 1879. The following year, his family moved to Munich, where his father and an uncle started a small electrochemical factory. After his father's business failed, in 1894 his family moved to Milan, Italy. Albert, then 15, was left behind in the care of relatives so that he could complete the requirements for the diploma he would need to enter a university.

Albert, however, paid increasingly less attention to his studies, until he was asked to leave the school. He then rejoined his family in Milan. In 1896, after passing a special examination, he was admitted to the Swiss Polytechnic Institute of Zurich in a 4-year program in physics and mathematics.

Einstein was soon bored by most of his courses at the Polytechnic. His insatiable appetite for reading seldom extended to the books required for his courses. Gifted with charm and wit, he made many friends, and attended cafes much more regularly than lectures. Among his new friends was classmate Marcel Grossman, a bright mathematics student. Having skipped so many classes, Einstein would have flunked his graduation examinations, if he had not been rescued by Grossman, who lent him his very careful notes. Einstein crammed for the examinations, and graduated in 1900.

In 1903, he married a former classmate at the Polytechnic; the following year, the first of their two sons was born. For two years, he got by with tutoring and part-time teaching, until his friend Marcel, through family connections, was able to get him an undemanding job at the Swiss Patent Office in Bern.

During the next three years, in the back room of his small apartment, Einstein conceived the revolutionary ideas that led to his prodigious outpouring of papers in 1905. At work, he quickly mastered his job of examining patent applications, and was able to bootleg time for his own research.

After his papers were published in 1905, the value of his work was soon recognized by some of the most distinguished scientists in Europe, particularly the German physicist Max Planck. This, however, did not catapult him to the international fame he would achieve later on. For a while, Einstein continued to work at the Patent Office, which he later called "that secular cloister where I hatched my most beautiful ideas".

In 1909, he left the Patent Office. Over the next few years, he moved to increasingly more attractive university positions, whose demands, however, interfered with his research. Finally, in 1914, at the insistence of Max Planck, he accepted a professorship at the University of Berlin, where he would be free to devote himself to his research.

After the outbreak of World War I, Einstein was an outspoken critic of German militarism. He continued, however, to be primarily absorbed in perfecting his General Theory of Relativity, on which he had been working for nearly ten years. This theory, which was published in 1916, was later described as "the greatest feat of human thinking about nature, the most amazing combination of philosophical penetration, physical intuition and mathematical skill". [2]

One of the predictions of the new theory was that light coming from a star, when passing by the Sun, would bend by some amount that Einstein had computed. This effect could be observed only during a solar eclipse. In spite of the war, news of Einstein's work spread quickly. A copy of his publication was submitted by a Dutch astronomer to the British Royal Astronomical Society, whose secretary was Arthur Eddington, a brilliant astrophysicist. Although there was much hostility among British scientists against German research, Eddington, who was a pacifist, embarked in the study of Einstein's work, in spite of the very difficult mathematics involved.

Eddington, then 34, had declared he would seek deferment from the draft as a conscientious objector. Instead of being sent to a labor camp, however, he was sent on an arduous trip to Africa to observe a full solar eclipse, which might confirm Einstein's theory.

In November 1919, the Royal Society of London announced that its scientific expedition to Africa had completed calculations that verified Einstein's prediction. Overnight, his name became a household word. Worldwide fame brought thousands of invitations to lecture and to write, most of which he turned down. For three years, however, he did travel widely to lecture on relativity, often arriving at his destination by third-class rail, with a violin tucked under his arm.

Although the 1920's were years of wide acclaim, Einstein concentrated on his new quest: to find the mathematical formulation of a "unified field theory" that would encompass both gravitation and electromagnetism, just as Maxwell had succeeded in unifying electricity and magnetism under a single theory. This would have been the first step, he felt, toward discovering the common laws that govern the behavior of everything in the universe, from electrons to planets. It turned out, however, to be a fruitless search, which occupied the rest of his life.

Einstein remained in Berlin until Hitler came to power in 1933. Although greatly admired by some Germans, he was disliked by the Nazis, who resented his Jewish heritage, outspoken pacifism and active support of Zionism. Soon after Hitler became Chancellor of Germany, Einstein renounced his German citizenship and left the country. He later accepted a full-time position at the newly founded Institute for Advanced Studies in Princeton, New Jersey.

He remained in Princeton for more than 20 years until he died. He lived with his second wife in a simple house, and most mornings walked to the Institute. In 1950, he published a new version of his unified field theory but, like the first version published in 1929, it was found untenable by most physicists. In comparison to his eminence a generation earlier, Einstein was now virtually neglected. In 1955, at the age of 76, he died in his sleep at Princeton Hospital.

THE SPECIAL THEORY OF RELATIVITY

Einstein's Postulates

In the years preceding his 1905 publications, Einstein had been largely out of the mainstream of physics. Working in the isolation of his small apartment in Bern and the "intellectual cloister" of his job at the Patent Office, he had not interacted with any professional physicists, which may have kept his imagination free of constraints. What guided his work was an unshakable faith in the unity of the universe and its laws.

Einstein was firmly convinced that the same laws of electromagnetism and optics "will be valid for all frames of reference for which the equations of mechanics hold good." He proposed that this conjecture be adopted as a postulate (something to be accepted on faith). Like Poincare's principle, Einstein's postulate states that the laws of both mechanics and electromagnetism are invariant to (do not change with) straight uniform motion without rotation. It represents a broader restatement of Newton's special principle of relativity, which applied only to the laws of mechanics. Einstein's theory of relativity is as much about the absolute nature of the laws of physics as it is about the relative nature of motion.

Einstein proposed then a second postulate: in empty space, light always propagates in all directions with the same speed c, regardless of whether its source or its observers are at rest or moving.

The two principles of the *invariance of the laws of physics* and the *constancy of the speed of light* could not be reconciled with Newtonian mechanics, which predicted that observers on different inertial frames would measure different values for the speed of light. Very boldly, however, Einstein adopted both principles as his fundamental premises, and then proceeded to show that their incompatibility with Newtonian mechanics could be resolved by drastically reevaluating notions about time and distance that had always been taken to be self-evident.

Gedanken Experiments

Einstein developed his theory with the aid of relatively simple mathematics. To explain his results, he used as examples what he called "gedanken experiments" (thought experiments), which are performed in one's mind, without too much concern for the practical details of implementation. In his examples, he used a train, which was then the fastest available means of locomotion. Today, he probably would have used a space ship.

In the "experiments" to be described, we will consider a phenomenally fast train moving at uniform speed on rails that are fixed to an inertial "platform" and go on and on, perfectly straight. The speed of this supertrain is some appreciable fraction of the speed of light.

There are two observers. One lives somewhere on the "ground" by the railway track. The second observer lives on the train, which is his “inertial platform." In order to keep these two observers clearly distinct in our minds, the observer on the ground will be called Stan as the one whose platform supposedly stands still. The observer on the train, instead, will be called Moe as the one whose platform is supposedly moving.

Actually, all we can say about the two observers is that their platforms are in relative motion with respect to one another. Stan claims that he and his whole platform are at rest, and that Moe and his train are moving at some speed s in the direction of the track. Moe, on the other hand, claims that he and the train are at rest, and that Stan and his whole platform, track included, are moving at the same speed s, but in the opposite direction. Neither can back up in any way his claim of being at rest.

Let us assume that, initially, everything is at rest. Stan and Moe know that there will be days when they will wake up and find themselves in relative motion, either because the train and Moe are moving, or because the platform, the track and Stan are moving. In anticipation of this, in order to conduct some experiments, they equip themselves with identical yardsticks, and with identical clocks.

They also build a hangar that fits the train exactly, see **Figure 12.1a**. At the two ends of the hangar and the train, they install two sets of special devices. One set will generate a momentary flash of light at the very instant the left ends of both hangar and train are in line; we will call this the L flash. Similarly, the other set will generate a momentary flash of light at the very instant the right ends of both hangar and train are in line; we will call this the R flash.

Length Contraction with Motion

In his work, Einstein stressed the fundamental role that the concept of simultaneity plays in the measurement of a length. Such a measurement must occur all at the same time; for instance, by matching simultaneously both ends of a rod to a ruler.

In Newtonian mechanics, it was considered self-evident that, regardless of their state of rest or motion, any number of observers would always agree on whether two events occurred at the same time, or which occurred first. They would also agree that the length of an object would not be affected by its being at rest or in motion. The following gedanken experiment shows that agreement on either point is not to be taken for granted.

Stan and Moe are now in relative motion with respect to one another. Stan claims the train is moving toward the hangar. Moe, instead, claims the hangar is moving toward the train. In any case, they want to compare the lengths of the train and the hangar in order to determine how length in the direction of motion is affected by motion.

One way they can do this on the fly is to take advantage of the L and R flashes. Moe places himself at the midpoint of his train; Stan, at the midpoint of his hangar (on the roof, safely out of the way). Using mirrors, each can see light flashes coming from either direction. Each is at a midpoint, and each agrees that the time it takes a light flash to reach him from one end of the train (or hangar) is exactly the same as from the other end. If the two flashes occur simultaneously, an observer at rest at exactly the midpoint will see the two flashes at the same time.

Keep in mind that each flash will propagate in all directions with a very high but still finite speed (186,000 miles per second). Neither observer will see a flash the very instant it is generated. When he sees it will depend on whether he is at rest, moving toward it, or away from it, and how far he is.

According to Einstein’s equations, Stan will see the flash from the left (L) before the flash from the right (R), whereas Moe will see R before L. Each is convinced the other is wrong. Each swears that he has rechecked with his yardstick the length of his train or hangar, and found it to be what it always was. Just what is going on? To follow their different points of view, see **Figures 12.1b, c and d**.

The disagreement between Stan and Moe as to which flash occurred first would not arise if they were not in relative motion. The disagreement would not arise either if light propagated instantly (with infinite speed in zero time). Since light propagates at a finite speed, two observers in relative motion will inevitably disagree about the sequence of events, and there is no way of determining which of the two is wrong. Each is right within his own frame of reference.

If two events occur at the same time *and* at the same place, they will be viewed as simultaneous in all reference frames. Disagreement about the simultaneity or sequence of two events can occur only when they are some distance apart. In this case, when large distances are involved, an observer can be present at best at only one of the events, and must obtain his information about the other event by means of some signal. Light is an extremely fast signal, but it too takes time to travel from one place to another.

Stan and Moe - who disagree on their measurements of lengths in the direction of their relative motion - are always in agreement when measuring lengths in a perpendicular direction, for instance, the height or the width of the train or the hangar.

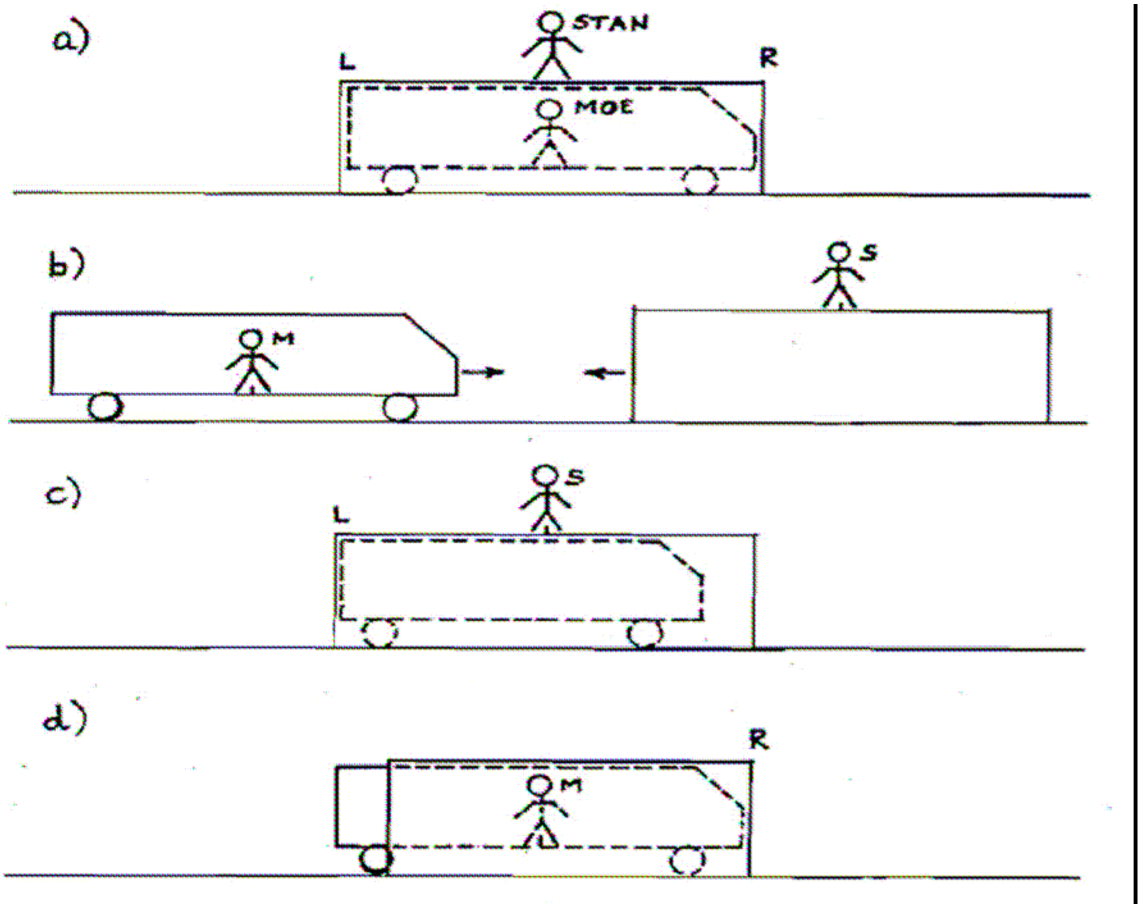

Figure 12.1 - Length contraction with motion

a) Stan and Moe both at rest. b) Stan and Moe in relative motion.

c) <u>Stan's viewpoint</u> - Stan and the hangar have not moved; Moe's train came speeding toward the left end of the hangar. Flash L occurred before flash R, which means that the train has become shorter. Once L occurred, Moe was moving toward R, which occurred second, and away from L, which occurred first. That is why he saw R before L. Since Moe claims he has rechecked the length of his train with his yardstick, it must mean that the moving train and everything on it, including the yardstick, have contracted (shrunken) in the direction of the motion, but Moe is unable to see the change.

d) <u>Moe's viewpoint</u> - Moe and the train have not moved; Stan's hangar came speeding toward the right end of the train. Flash R occurred before flash L, which means that the hangar has become shorter. Once R occurred, Stan was moving toward L, which occurred second, and away from R, which occurred first. That is why he saw L before R. Since Stan claims he has rechecked the length of his hangar with his yardstick, it must mean that everything on Stan's moving platform, including the yardstick and the hangar, has contracted in the direction of motion, but Stan is unable to see the change.

Time Dilation

Two observers in relative motion will disagree not only on the sequence of events, but also on the very rate at which time flows! This will be shown by another gedanken experiment performed by Stan and Moe using two identical special clocks. See **Figure 12.2.**

Stan makes the startling discovery that Moe's clock, just by virtue of the fact that it is moving, runs slower than his, and that, the greater the speed of the train, the greater the discrepancy between the rhythms of the two clocks. By the same reasoning, however, Moe - who believes that he is at rest, and that Stan is moving - has concluded that Stan's clock runs slower than his.

If, at this point, you feel totally confused, and ready to dismiss the whole thing as nonsensical, you will join a very large crowd of people who have had the same reaction. It is a fact, however, that the theory of special relativity has been verified countless times at speeds close to the speed of light.

You might wonder whether the strange conclusions reached about time have something to do with the peculiar light clocks that have been considered. Keep in mind, however, that one of Einstein's two articles of faith states that the laws of physics are invariant from one inertial frame to another, which requires that it be impossible to detect the motion of one's own inertial frame of reference.

Suppose now that, in addition to the light clocks, Stan and Moe are given two stately grandfather clocks, or whatever you would consider to be "real" reliable clocks. If a discrepancy were found between a light clock and a "real" clock, it would be an indication of motion. But this would violate the very article of faith that was accepted at the beginning. If anything at all changed because of straight uniform motion, one would be able to tell he was moving. If we accept Einstein's postulates, we must agree that the "real" clocks will behave just like the light clocks.

In the words of Nobel Laureate Richard Feynman, "If all moving clocks run slower, if no way of measuring time gives anything but a slower rate, we shall just have to say, in a certain sense, that 'time itself' appears to be slower in our supertrain. All the phenomena there - the man's pulse rate, his thought processes, the time it takes to light a cigar, how long it takes to grow up and get old - all these things must be slowed down in the same proportion. Otherwise, he would be able to tell that he is moving."[1]

Although two observers in relative motion disagree on their measurements of time and distance, they agree on the value they measure for the speed of light. Each, however, claims that the other uses the wrong way to arrive at the correct result.

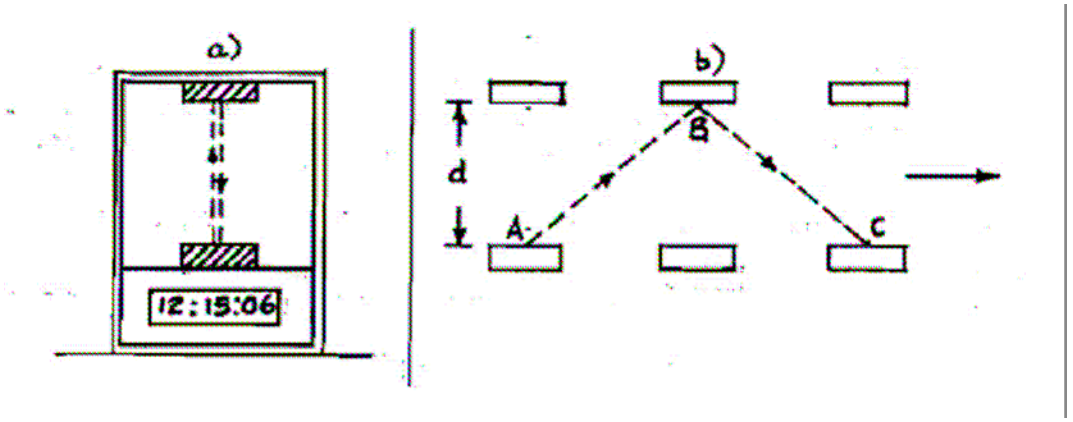

Figure 12.2 - A light clock - Time dilation with motion

The "light clock" shown in a) above consists of an upper and a lower element separated by a fixed distance. The lower element generates a short pulse of light, whose waves propagate in all directions. When the pulse reaches the upper element, which is a plain mirror, it is reflected back toward the lower element. Here, a new pulse is emitted the same instant the old one is received. Also, some kind of a "click" is generated that can be counted to accumulate time into seconds, minutes and so on. Basically, this is no different from an ordinary clock, which counts the time intervals between consecutive swings of a pendulum, or consecutive oscillations of a balance wheel - time intervals that are believed to be always identical.

Stan and Moe have mounted their identical clocks in an upright position, perpendicularly to their relative motion. When they are not in relative motion, they have satisfied themselves that the two clocks keep exactly the same time. When they are in relative motion, each clock looks perfectly normal to its owner.

Stan, who believes to be at rest, notices something peculiar about Moe's clock, see b) above. Since the train is moving, by the time a light pulse reaches the upper element of the clock, the latter has moved somewhat. The lower element has moved even further by the time the reflected pulse gets back to it. Instead of traveling the straight up and straight down distance 2 x d, the pulse has traveled the longer distance ABC. The time interval between "clicks" has dilated (become longer), which means that Moe's clock is running slow. Its readings (say, in seconds) are smaller than they should be. For Moe, however, it is Stan's clock that runs slower!

The "Gamma" Correction Factor

In a general way, we have seen how, according to Einstein, Stan and Moe differ in their assessments of time and distance. Each sees the other make measurements of time that are too small, and measurements of distance (in the direction of relative motion) that are too large. Let us see what the quantitative extent of the discrepancies is.

Time Measurements

If Moe says that some interval of time is, say, 100 seconds according to his clock, Stan "knows" that he must use some correction factor, say 2: the "correct" number is 200 seconds (according to Stan's clock). We will call this correction factor "gamma". Einstein concluded that "gamma" depends on the ratio of the speed of relative motion s divided by the speed of light c, or s/c.

If s is much smaller than c, "gamma" is only very slightly larger than 1. For instance, if our supertrain moved at the speed of the Earth around the Sun (about 19 miles/sec.), s would be only one tenthousandth (or 1% of 1%) of c, and gamma would be 1.000000001. It would exceed 1 by only one part in 100 million. There would be essentially no disagreement between Stan and Moe. That is why, at ordinary speeds, we do not notice any relativistic effects.

Only when s is an appreciable percentage of c does the correction factor become significant. As s gets closer and closer to c, "gamma" gets larger and larger, growing faster and faster. For instance, "gamma" is 1.25 when the ratio s/c is 60%; 2.29 when the ratio is 90%; 7.09 when the ratio is 99%. It jumps to 5000 when the ratio s/c is 99.99%; it becomes infinite when s is equal to c.

Length Measurements

Using his yardstick, Moe measures some object on his train to be, say, 20 inches long in the direction of relative motion. Stan, however, "knows" that both the object and the yardstick have shrunken, and that to get the correct number of "real" inches, he must *decrease* Moe's number by some correction factor. It turns out that this correction factor is the same "gamma" just discussed, except that now we must divide by it. For our example, Stan claims that the "correct" length is 20/2=10 inches (according to Stan's yardstick).

Paradoxically, Moe makes exactly the same kind of corrections, using the same gamma factor. Moe believes that he is at rest and Stan is moving. Stan's clock is the moving clock that runs slow, and Stan's yardstick is the moving yardstick that has shrunken.

Relativistic Mass

Einstein's faith in his two postulates led him to revise drastically the concepts of time and distance. Since these concepts are the very foundation of how we describe motion, the whole science of mechanics would appear now to rest on very shaky ground. And yet, for more than two hundred years, Newton's equations had yielded results in remarkable agreement with the observed motions of the planets.

How could this be? The answer is that Newton and his successors had dealt only with a range of phenomena in which the speeds involved were much smaller than the speed of light.

The changes that need to be made in Newtonian mechanics center on the concept of mass. In the Principia, Newton defined mass as the "quantity of matter" contained in a body. He assumed it was always constant, whether the body was at rest or in motion.

Einstein tells us, instead, that the mass of a body is not constant but increases with speed. If a given body is at rest with respect to some observer, the latter will measure a certain mass, which is called the "rest mass" of the body. This is the value of mass that Newton would have used.

If the body, however, is moving at some speed s with respect to the same observer, the latter will measure a larger mass equal to the "rest mass" times the now familiar correction factor "gamma".

You will recall that, as long as the speed s is much smaller than the speed of light c, "gamma" is for all practical purposes equal to 1. Mass in motion and rest mass are then essentially identical. Under these conditions, we can disregard the correction factor introduced by relativity, and use Newton's equations with very satisfactory results. As the speed s becomes a sizeable fraction of c, however, the gamma factor increases more and more: the mass of the moving body offers an ever increasing inertia.

The Cosmic Speed Limit

In Newton's mechanics, since mass is constant, if we want a given body to increase its speed, say, by one foot per second in one second, it does not matter whether the body is starting from 100 or 100,000 feet per second. For a given mass, the force to be applied depends only on the acceleration, the change in speed per unit time.

This is not so, however, in Einstein's mechanics. The nearer a (relative) speed is to that of light, the more difficult it becomes to increase it further. A body resists a change more strongly not only when its rest mass is greater, but also when its speed is greater.

If a force is applied to a body, both the mass and the speed of the body will increase, but not at the same rate. At slower speeds, mass increases more slowly than speed; at higher speeds, mass increases much faster than speed. The speed of the body never exceeds the speed of light c, but gets ever closer to it.

When the speed s is equal to c, the mass becomes infinite: no further acceleration is possible. Thus, the speed of light c represents a cosmic speed limit, whose enforcement is automatically assured by the way mass increases with speed. No object can be accelerated beyond the speed of light.

Mass-Energy Equivalence

According to Newtonian mechanics, if we apply a constant force to a particle, the particle keeps picking up speed at a constant rate, increasing its kinetic energy. This kinetic energy is "paid" for by the "work" we do by applying the force. This is an adequate description of things as long as the speed of the particle is much smaller than the speed of light.

If we apply a force to a particle moving at very high speed, the speed increases at a slower and slower rate. Since we are still doing "work", the particle's energy must be increasing, but less and less of it is going into kinetic energy. Where is the rest of the energy we are "paying" for with our "work"? Since the mass of the particle is also increasing at the same time, we might suspect that the mass itself represents energy. This indeed turns out to be the case. Einstein showed that the total energy of the particle is

$$\text{Energy} = \text{mass x the square of the speed of light} = m \times c^2$$

The theory of relativity leads us to conclude that the mass of a body is a measure of its energy content. Energy, whatever its form, has mass. Mass can be converted into energy, and vice versa. Note that, when multiplied by the square of the speed of light (a very large number), even a small mass can yield an enormous amount of energy.

Since the atomic bombing of Hiroshima, Einstein's famous equation has been popularly associated with nuclear energy. Actually, Einstein had no direct involvement in the development of nuclear energy for either war or peace. His equation applies to all forms of energy. In nuclear reactions, in which mass is partially converted into energy, we can have mass changes of about one part in 1000. On the other hand, heating water from its freezing point to its boiling point increases its mass by merely one part in 100 billions.

Before relativity, physics recognized two fundamental laws of conservation, which appeared to be independent of one another: the conservation of energy and the conservation of mass. The theory of relativity united them into a single law: the conservation of mass-energy.

SPECIAL RELATIVITY IN A NUTSHELL

In his theory of special relativity, Einstein reached the following major conclusions:

- Maxwell's equations, which were at first under suspicion, actually required no change.
- It was Newton's laws of motion that needed to be changed because of a

new definition of mass, no longer viewed as a constant, but as increasing with speed.

- Maxwell's original equations and Newton's modified equations are both invariant from one inertial platform to another.
- Measurements of time and distance are relative to the observer's platform. To somebody "at rest" looking on somebody else's "moving" platform, time there appears to flow more slowly, and distances appear to shrink in the direction of relative motion. Events that are viewed on one platform as simultaneous, may not appear so in another platform. An event that is seen on one platform to occur *before* some other event, may be seen to occur *after* on another platform.
- No object can be accelerated past the speed of light.
- Mass and energy are equivalent.

We considered before the question: If a train is moving away from you at 20 miles/hour and a jogger on the train is running at 5 miles/hour, also away from you, what is the speed of the jogger relative to you? Newtonian mechanics told us: "Obviously, 25 miles/hour, the sum of the two speeds."

This simple way of combining speeds is no longer valid in general. It is still essentially valid for speeds that are small relative to the speed of light, but not for higher speeds. When two speeds are combined by Einstein's rules, even if either or both are equal to c, we still get no more than c! The Michelson-Morley experiment failed to show a winner because it was based on the wrong assumption that speeds could be combined by addition or subtraction.

It should be noted again that all aspects of special relativity have been repeatedly confirmed. Physicists studying subatomic particles work daily with objects traveling at speeds close to the speed of light; their experiments confirm Einstein's predictions. At ordinary speeds, on the other hand, the effects of relativity are totally negligible.

Since special relativity, space and time are viewed not as two independent entities, but as a single 4-dimensional entity, "space-time", something that our three-dimensionally oriented brain finds most difficult, if not impossible, to visualize.

Chapter 13

GENERAL RELATIVITY

The General Principle of Relativity

Special Relativity stemmed from the work of several people. It was motivated by the definite need to reconcile Newton's and Maxwell's laws. The General Theory of Relativity, instead, rests almost exclusively on the insight of one man, Einstein, who was determined to pursue to the limit the implications of a philosophical principle.

With his Special Theory of Relativity, Einstein had shown that both Maxwell's equations and Newton's modified equations of motion are valid regardless of the particular inertial frame from which one happens to observe the universe. As a result, neither Stan nor Moe could detect being in uniform motion.

Einstein embarked now on an even bolder venture: to show that the laws of physics can be stated in a form that is equally valid not only for *inertial* frames of reference (which move with straight uniform motion), but also for *any* frame of reference, *whatever* its motion (for instance, linearly accelerating frames, or rotating frames). Einstein saw no reason why inertial frames should have a privileged status. It is the same universe, he felt, whatever the reference frame we observe it from.

This is the philosophical principle that motivated Einstein: it is called the *General Principle of Relativity*. It implies, however, that it should be impossible to detect any motion of one's own frame of reference - uniform or otherwise. Einstein's venture seemed doomed from the start because of the well familiar jerking effect that accompanies any change in speed and/or direction of motion. There appears to be something absolute about acceleration. A reevaluation of gravity, however, gave a way out.

To explain the motions of bodies, earthly as well as heavenly, Newton had proposed a gravitational force of attraction which, instantly, even across enormous distances, can act on two bodies. Instantaneous action at a distance, however, is incompatible with the Special Theory of Relativity, which asserts that no physical influence can travel faster than the speed of light. Inspired by Maxwell's theory of electromagnetic fields, instead of a force of gravity, Einstein postulated a gravitational field that propagates at the speed of light.

Gravitational vs. Inertial Mass

There was something about Newton's force of gravity that had puzzled physicists from the start. The term "mass" that appears in Newton's law of gravitational force could be replaced by the more specific term "gravitational mass". Similarly, the term "charge" that appears in Coulomb's law of electrostatic force could be replaced by the term "electrical mass". Finally, in Newton's law of motion (force = mass x acceleration), the term "mass" could be replaced by the more specific term "inertial mass".

Whether the force applied to a body is from a gravitational or an electrical source, it is the *inertial* mass of the body that determines the resulting acceleration. There did not seem to be any reason why either the gravitational or the electrical mass should equal the inertial mass, and yet the gravitational mass does.

In Newtonian mechanics, this equivalence of inertial and gravitational mass had been regarded as just a strange coincidence. Einstein was the first to recognize that this "coincidence" indicated something peculiar about the nature of the gravitational force. His interpretation of the situation led to the second fundamental principle of his General Theory of Relativity: the Principle of Equivalence.

The Principle of Equivalence

As previously noted, Newton's laws of motion are valid only with respect to an inertial frame of reference. To achieve the ideal conditions required for a truly inertial frame, we need to place ourselves in some region of outer space so far removed from stars and other appreciable masses that we can assume no gravity force that would cause acceleration.

Let's consider now another of Einstein's thought experiments. This time, instead of a train, we have a very special elevator, which he described as "a spacious chest resembling a [windowless] room with an observer inside who is equipped with [measuring] apparatus". We will call this observer Eliot as a way of remembering that he is the guy in the elevator. Let us imagine this elevator in the depths of intergalactic space far from all matter, where we will also place our friend Stan and his inertial reference frame.

If the elevator were drifting in empty space with constant velocity, it too would be an inertial reference frame. Stan and Eliot would disagree about each other's clocks and yardsticks, but not about their interpretation of physical phenomena.

Let us assume, instead, that somebody is pulling on a rope attached to a hook in the middle of the elevator's roof, applying a constant force. As a result, the elevator is moving "upwards" with uniformly accelerated motion. We have now a *non-inertial* reference frame. With the right force, we can cause a constant acceleration equal to 1g, i.e., equal to Galileo's acceleration of gravity near the Earth's surface (about 32 feet per second every second).

In his elevator accelerating at 1g, Eliot feels his normal weight, as he can verify on a scale that is part of his measuring equipment. In fact, everything feels quite normal. If he drops an apple he is holding in his hand, he will see it fall with an acceleration g. Even if he drops a much heavier object, he sees it too fall with exactly the same acceleration. If he drops both objects at the same time, both will hit the floor at the same time. From all that he can see inside his windowless laboratory, he concludes that his elevator is at rest in a gravitational field just like Earth's.

From his inertial reference frame, Stan gives a totally different interpretation of things. The elevator is moving upwards with uniformly accelerated motion, pushing on Eliot's feet and carrying him with it. As long as Eliot holds on to the apple, the apple is accelerating along. When Eliot lets go of the apple, he stops applying a force. By Newton's law of inertia, the apple continues to move upwards at constant speed, but soon enough the floor of the accelerating elevator catches up with it and starts pushing it along. A second object released at the same time will continue to move upward at the same uniform speed, until it too is hit by the floor at exactly the same time as the apple. Eliot does not realize what is really happening - he has invented a fictitious force of gravity to explain things and retain his illusion of being at rest.

This dual interpretation of the same phenomena is made possible by the fact that gravitational mass and inertial mass are equal. If, for instance, the inertial mass of a body were twice its gravitational mass, in his no-gravity environment, Eliot would measure twice his normal weight. He would be able to tell that things are different.

We should not conclude, however, that the existence of a gravitational field is always only an apparent one. Inside an elevator that is accelerating in empty space, the forces of pseudo gravity are all equal and parallel to the direction of motion. Inside an elevator standing on the surface of the Earth, the forces of gravity vary from point to point, and are all converging toward the center of the Earth. The differences, however, are imperceptible within the small space of the elevator.

These considerations led Einstein to state his Principle of Equivalence as follows:

Within a small space, there is no measurable difference between the effects of a gravitational field, and the effects that result from the acceleration of one's reference frame.

The Principle of Equivalence tells us that what Eliot observes in his elevator accelerating in no-gravity intergalactic space, is the same as what he would observe if his elevator were at rest in a gravity field. From this principle, Einstein derived important insights that guided him in the development of a mathematical theory.

If a pulse of light were to enter at some point near the top of the accelerating elevator and proceed in a direction perpendicular to the motion of the elevator, it would appear to Eliot to follow some downward curved path and leave the elevator at some point lower than the point of entry. Using the Equivalence Principle, Einstein could predict then that light would be bent in the vicinity of a large mass such as the Sun.

Einstein predicted also that time is affected by a gravity field. He imagined Eliot experimenting with two identical clocks: one near the bottom of the elevator, the other near the top. Einstein showed that Eliot would conclude that the clock on top runs faster than the clock on the bottom.

Note that, when we were following the antics of Stan and Moe, we were comparing identical clocks on separate inertial frames in relative motion. Stan and Moe were claiming that each other's clock ran slower. Now we are talking of two identical clocks at rest on the *same* non-inertial frame, and, according to Eliot, the top clock runs faster than the bottom clock.

Applying the Principle of Equivalence, Einstein could predict that, in the presence of a gravity field, a clock at a higher level would run faster than a clock at a lower level. This has been verified experimentally. In the 1960's, for instance, researchers found that clocks in the mile-high city of Boulder, Colorado, gained about 15 billionths of a second a day compared with clocks near sea level in Washington, D.C. Near a massive star, time dilation would be much more pronounced.

Special Relativity told us that time is affected by motion. Now, General Relativity tells us that time is also affected by gravity. More speed slows down a clock, and so does more gravity.

"Help me, Marcel, or I'll Go Crazy!"

Confusing as things might have been with special relativity, at least everything made sense on any particular inertial "platform", where clocks and yardsticks gave consistent measurements. Two people on the same platform could agree on measurements of time and distances. It was only with somebody on another inertial frame that disagreements arose.

When Einstein started considering non-inertial frames moving in various ways, he faced a totally bewildering situation. It is no longer possible to define a single length or a single time that has the same meaning everywhere, even within a single reference frame. From one point to another, a yardstick shrinks or expands, a clock runs faster or slower. Add the fact that the mass of a body changes with speed, and you can appreciate the overwhelming difficulties Einstein was confronted with, while trying to develop a mathematical theory of gravity that would hold in any frame of reference. The problem was stretching his mathematical skills to the limit. Once again, rescue came from his friend Marcel Grossman, to whom he wrote in desperation, "Help me, Marcel, or I'll go crazy!"

Marcel, who was by now a professor of mathematics, introduced Einstein to the work of the German mathematician Bernhard Riemann, who, without any particular application in mind, had developed just the mathematical tools Einstein needed.

Riemann's work was about "non-Euclidean" geometry, the geometry of curved surfaces and spaces. This field was extensively studied by Riemann and others during the 19th century. Until then, the only geometry known was the one built upon Euclid's Elements, which was considered the "true" geometry of space.

If, instead of a plane, we consider some other type of surface, say, a sphere, the rules of geometry change completely. On a plane, for instance, the angles of any triangle always add to 180 degrees. The angles of a triangle drawn on a sphere, instead, always add to more than 180 degrees. A triangle drawn on a sphere can have each of its angles equal to 90 degrees.

On any kind of surface - spherical, pear-shaped or whatever - given two points A and B, there is a path of shortest distance between them, which is called the "geodesic" between the two points. On a plane, a geodesic is a segment of a straight line. On the surface of a sphere, a geodesic is an arc of a circle, whose center coincides with the center of the sphere. Each type of surface has its own geometry and its own kind of geodesic.

Einstein's Theory of Gravity

Einstein was interested in a geometry of curved space because he speculated that the presence of matter causes the warping of a 4-dimensional space-time, whose geometry becomes non-Euclidean. The gravitational interaction between two bodies arises from the geometry of space itself.

A helpful 2-dimensional analogy is to think of space-time as if it were a huge very flexible rubber sheet that is stretched tightly across a frame, like a cosmic trampoline. Imagine now massive bodies (such as stars) placed here and there on the trampoline. Pressing down on the surface of the rubber sheet, each star will warp it causing a depression, called a "gravity well". (You may have seen a rigid model of a gravity well in a science museum, where you can release a coin or a small steel ball, and watch it go around and around down to the bottom of the well.) Smaller masses will have a smaller warping effect. Where there is little significant mass, the surface is essentially flat.

A planet orbiting around a star moves the way it does on the cosmic trampoline simply because it is following the warped contour of the surface, not because some Newtonian force of gravity is acting on it from a distance.

A comet moving toward a star will descend into the star's gravity well. It may remain trapped in the well, orbiting along its sides forever. If the comet is fast enough, however, instead of being trapped, it will emerge from the well and continue its journey, somewhat deflected from its original path.

Where there is no mass, the geometry of space is Euclidean and Newton's law of inertia holds. Here, without the intervention of any force, a body moves on its own with constant speed along a *straight* line. Einstein generalized Newton's law of inertia to state that a body moves on its own along a *geodesic* line. Within a gravitational field, however, geodesic lines are no longer straight lines, but are still paths of minimum distance.

For Newton, forceless motion by inertia was natural and did not require an explanation. For Einstein, a body falling or orbiting under the influence of gravity is also in a forceless natural state of motion, but in a warped space-time, in which it moves along a geodesic line.

In Newton's theory, space and time were distinct, independent entities. As a universal clock ticked away, space was a passive, immutable stage on which the phenomena of physics unfolded without either affecting, or being affected by, the stage.

Special relativity combined space and time into a single entity, space-time, which became the new passive, immutable stage of physics. With general relativity, instead, space-time becomes an integral part of the whole process. The presence of matter causes space-time to warp; the warping of space-time, in turn, affects the motion of matter. The stage and the play mutually affect one another. As the eminent physicist John Wheeler put it very succinctly, matter tells space how to curve, and space tells matter how to move.

Confirmations of General Relativity

Only in the presence of huge masses can we expect any measurable differences between the predictions of general relativity and those of Newton's gravitational theory.

Consequently, most experimental tests of general relativity have been based upon astronomical observations. As previously mentioned, Eddington's expedition to Africa provided confirmation of Einstein's prediction that light too travels along geodesic lines and, therefore, will be bent in the vicinity of a large mass such as the Sun.

Another confirmation of the theory was provided by Einstein's ability to account for a peculiarity in the orbit of Mercury, the planet that is closest to the Sun. According to Kepler, a planet should retrace the same elliptical orbit again and again, always coming back to the same "perihelion", the point on its orbit closest to the Sun. In the case of Mercury, evidence accumulated during the 19th century revealed a discrepancy that could not be completely explained in Newtonian terms. Instead of remaining stationary with respect to the stars, Mercury's elliptical orbit as a whole slowly rotates.

What is observed is a small advance of Mercury's perihelion by 543 seconds of arc (about one seventh of a degree) per century. Taking into consideration the gravitational pull of the Sun and of all the known planets, Newtonian mechanics could account only for an advance of about 500 seconds of arc per century. Einstein's theory can account for the remaining 43 seconds. For the orbits of the other planets, Einstein's theory gives the same results as Newton's.

The general theory of relativity is widely accepted as the most satisfactory explanation for the gravitational interaction. It has achieved its most dramatic success in the field of cosmology, the study of the universe as a whole.

Chapter 14

INSIDE THE ATOM

In the late 1800's, a series of discoveries led to a new understanding of the atom, which was found to have a complex internal structure.

RADIOACTIVITY

Becquerel

In 1896, the French physicist Henri Becquerel discovered that crystals of a uranium compound emitted a mysterious form of penetrating radiation. Later, he traced this radiation to the uranium atoms in the crystals. Later still, this ability to emit radiation spontaneously was called "radioactivity".

The Curies

Becquerel's discovery inspired a young physicist, Marie Sklodowska Curie (1867-1934), to devote her life to the study of radioactivity. Born in Poland, at 24 she moved to Paris. There, she studied mathematics and physics, and married another physicist, Pierre Curie. Together, they began investigating the radioactivity of certain chemical elements. In 1898, the Curies discovered two strongly radioactive elements, polonium (which Marie named in honor of her native country) and, a few months later, radium.

Overcoming enormous difficulties in relatively primitive laboratory conditions, Marie and Pierre succeeded in isolating radium in pure form, an undertaking that would have required industrial resources. In 1903, they shared with Becquerel the Nobel Prize for physics.

In 1906, Pierre died at 46, killed by a horse-drawn wagon while crossing a street in bad weather. Marie was unanimously chosen to succeed him in his professorship, the first time a woman had ever been elected to teach at the prestigious Sorbonne, the University of Paris.

In 1911, she received the Nobel Prize for chemistry, becoming the first person to be awarded two Nobel Prizes. (In 1930, her daughter Irene would also receive a Nobel Prize for her work in radioactivity.) Marie Curie died in 1934 at the age of 67, from leukemia caused by excessive exposure to radiation.

Rutherford

The connection between radioactivity and the very structure of matter was revealed by the work of Ernest Rutherford (1871-1937), one of the greatest experimental physicists of all times. He was born in New Zealand, in a family of modest means. In 1895, after graduating from college, he left New Zealand with a 2-year scholarship to study at Cambridge, England. Here, he began to work as a research assistant to J.J. Thomson, who would later discover the electron.

Rutherford showed that the Becquerel rays consisted of two distinct types of radiation: the "alpha" rays, and the much more penetrating "beta" rays. Later, he showed that the alpha rays are positively charged helium atoms stripped of their electrons, whereas the beta rays are electrons.

(A third type of radiation, the gamma rays, was discovered in 1900. Unlike the alpha and beta rays, which are particles, the gamma rays are electromagnetic waves. It is because of the more penetrating beta and gamma rays that radiation has acquired its frightful reputation. Even a sheet of cardboard, on the other hand, is sufficient to stop a beam of alpha particles.)

In its early years, radioactivity was a most mysterious phenomenon. It seemed to defy the principle of energy conservation. The radiation from radioactive substances carried enormous energy, which exceeded by far that produced in any chemical reaction.

In 1898, Rutherford left Cambridge to accept a professorship at McGill University in Canada. Here he collaborated with Frederick Soddy, a talented young chemist. In 1902, they concluded that, as the result of radioactivity, atoms of one element spontaneously disintegrate into atoms of an entirely different element.

This "transformation" theory was opposed by many chemists, who strongly believed in the immutability of matter. To them, the idea that atoms of one element could break down into atoms of another element was a throwback to the medieval alchemists, who tried to transform lead into gold.

In 1908, Rutherford moved back to England to teach at Manchester University. Here, in 1911 he made his greatest contribution to science with his nuclear theory of the atom, as discussed in a later section. In 1919, he succeeded his former mentor, J.J. Thomson, as director of the prestigious Cavendish Laboratory.

Rutherford received many honors in recognition of his accomplishments. In 1908, for his work in radioactivity, he was awarded the Nobel Prize for chemistry (a choice he found disappointing, since he considered himself a physicist). He was knighted in 1914, and made Baron Rutherford of Nelson in 1931. When he died in 1937, he was buried in Westminster Abbey.

THE ELECTRON

For 25 centuries, atoms had been conceived as the ultimate indivisible building blocks of matter. By 1900, however, there was strong evidence to the contrary.

It is now a basic assumption of physics that the properties of matter depend on the properties and interactions of sub-atomic fundamental particles. Of these, the first to be discovered was the electron in 1897.

In the late 18th and early 19th century, it was popular to think of an (electrically) charged object as if it contained some sort of a fluid that could easily flow through a metal, but not through an insulator. Charge was considered to be continuous, that is, it could be increased or decreased by infinitesimal amounts.

A different theory of electrical charge began to evolve toward the middle of the 19th century, in order to explain consistently how current is conducted in metals, liquids and gases. According to this theory, charge is quantized, that is, it can be increased or decreased only in multiples of a single indivisible quantity. This fundamental unit of charge was called the "electron"; it is traditionally represented by the letter e.

In the 1890's, many physicists were experimenting with the flow of electric current through various types of sealed glass tubes, from which air had been evacuated as much as the technology of the day permitted. Sealed inside such a vacuum tube were a heated filament (the "cathode") and another terminal (the "anode"), with an appropriate voltage applied across the two.

The small amount of gas still in the tube could be seen to glow, revealing the paths of negative charges flowing from the heated cathode to the positive anode. At the end of the 19th century, there was intense speculation about the mysterious nature of these paths, which were called "cathode rays".

In 1897, J.J. Thomson (1856-1940) showed experimentally that these cathode rays marked the paths of material "corpuscles" (tiny bodies), each carrying a negative charge. In a specially designed vacuum tube, the cathode rays could be sharpened into a narrow beam, which, upon striking a specially coated screen, caused a phosphorescent spot. (We have here the basic idea for the present-day cathode ray tubes, or CRT's, used in TV sets and computer monitors.)

Thomson showed that the narrow beam behaved just as one would expect for a stream of negatively charged particles under the influence of electric and magnetic fields. Although he could not determine either the mass or the charge of a corpuscle, he was able to compute their ratio.

In 1897, Thomson announced the discovery of a new form of matter, the cathode-ray corpuscles, which were called the "atoms of electricity". Later, they were called electrons after the fundamental unit of negative charge, of which they are the carriers.

Between 1909 and 1911, the American physicist Robert A. Millikan performed a series of experiments which proved directly that charge is quantized, and yielded the actual value of the fundamental unit of charge. Although elementary particles of various types differ significantly in mass, only three values have ever been observed for their charge: +e, -e, and 0 (for neutral particles).

From Millikan's fundamental charge and Thomson's mass-to-charge ratio, it became possible to compute the incredibly small mass of an electron. The combined mass of a billion billion billion electrons is only one gram!

Thomson's electron was immediately seen as a building block in the structure of atoms. Since matter is electrically neutral, the individual atoms that make up matter would have to be neutral themselves. As indicated by a variety of experiments, the heavier the atom, the larger the number of electrons contained in it. Their combined negative charge would have to be neutralized by an equal but positive charge in the rest of the atom.

The simplest atom, that of hydrogen, would contain a single electron carrying one unit of negative charge, and an equal but positive charge, with which would be associated a mass about 1800 times the mass of the electron. Thus, practically the entire mass of the hydrogen atom would be associated with its positive charge.

There was no direct evidence, however, on how the electrical material might be arranged inside an atom. In 1902, Lord Kelvin proposed an atomic model in which the positive charge was distributed throughout a very small region, possibly a sphere, within which the electrons were embedded like raisins in a cake.

In recognition of his work, Thomson was awarded the Nobel Prize for Physics in 1906, and was knighted in 1908. An outstanding teacher, his contributions to physics rest almost as much on the work he inspired in others, as on his own. Seven Nobel prizes were awarded to physicists who worked under him.

THE NUCLEUS

Rutherford's work in radioactivity had convinced him that the structure of the atom could be studied by shooting a narrow beam of alpha particles at a very thin foil of metal. Careful measurements of how the alpha particles were scattered (deflected) might give a clue about the structure of the atoms bombarded by the particles.

This was the basic idea behind a series of experiments that were performed in 1909 at Manchester University by one of Rutherford's collaborators, Hans Geiger, who invented the first detector of individual alpha particles (the Geiger counter). He found that alpha particles going through thin foils of metal were deflected by an angle that was usually small, in the order of one degree.

Rutherford suggested investigating whether any alpha particle could be scattered by a large angle. He did not expect that it could: after all, an alpha particle was a (relatively) massive particle traveling at about 10,000 miles per second. The probability of a particle being scattered backward could be computed to be very small. Anyway, it seemed to be a good project to further the training of Geiger's young assistant, Ernest Marsden.

A special apparatus was built for the experiment. In it, inside a vacuum chamber, a radioactive source emitted alpha particles, which were channeled into a narrow beam aimed at a very thin gold foil suspended at 90 degrees. There were detectors to count alpha particles that might be deflected at various angles.

A few days later, with great excitement, Geiger reported to Rutherford that some of the alpha particles had been deflected backwards by substantial angles. Rutherford was astonished. Later, he wrote "It was quite the most incredible thing that has ever happened to me in my life. It was almost as incredible as if you fired a 15-inch [artillery] shell at a piece of tissue paper and it came back and hit you."

The results of the Geiger-Marsden experiment were not consistent with the raisin-in-the-cake model of the atom. From his calculations, Rutherford concluded that the atom had to have the great bulk of its mass and all of its positive charge concentrated in a sphere much smaller in radius than the whole atom. The atom was mostly empty space! Objects that appear to be solid are actually made of these empty spaces, and are held together by forces acting on electric charges.

If we enlarged the atom to a sphere 520 feet in diameter, the nucleus sitting at the center would be a sphere with a diameter of only one sixteenth of an inch (100,000 times smaller): the electrons would be mere dust specks somewhere between the two spheres.

In 1911, instead of the raisins-in-the-cake model, Rutherford proposed a new model, which has become the foundation of our present understanding of atomic structure. In this model, an atom consists of a highly concentrated positive charge (the nucleus) and a number of electrons spread out somehow over a much larger region.

The results of the Geiger-Marsden experiment could now be explained as follows. Since an atom is mostly empty space, the alpha particles rarely came near a nucleus. Little or no deflection was experienced by the great majority of particles. The closer an alpha particle flew by a nucleus, the larger the angle of deflection due to the repulsive force between their positive charges. The few very large deflections observed were due to near misses.

INSIDE THE NUCLEUS

Later, it was discovered that the nucleus itself has an internal structure. At first, it was believed that nuclei consisted only of one or more "protons", each with a positive charge equal and opposite to the charge of an electron, and a mass about 1800 times that of an electron. In 1932, the English physicist James

Chadwick discovered another particle very similar to the proton; since it has no electric charge, it was called the "neutron".

Except for the simplest form of hydrogen with one proton and one electron, all stable nuclei contain at least as many neutrons as protons. All atoms contain equal numbers of protons and electrons, and are therefore neutral. It is the number of electrons in an atom that determines the chemical element of the atom (i.e., whether it is gold or lead) and its chemical properties. It is called the "atomic number". In the periodic table (see end of Chapter 5), the elements are arranged in increasing order of atomic numbers. As we move from one element to the next, we find one more electron and one more proton in the atom.

Elements can come in varieties, called "isotopes", which differ in the number of neutrons in the nucleus. For instance, carbon with 6 protons has two isotopes: one with 6 neutrons, and one with 7. The number of "nucleons" (protons + neutrons) is called the "mass number". One isotope of Uranium, for instance, with 92 protons and 143 neutrons, has a mass number of 235; it is accordingly designated as Uranium-235.

The largest number of protons in any naturally occurring nucleus is 92, in uranium. Physicists have been able to manufacture nuclei with up to 106 protons, but they are unstable. A complex nucleus may be compared to a wobbly drop of water that hangs from a faucet with a tendency to split into smaller droplets. Radioactivity occurs when an unstable complex nucleus spontaneously "transmutes" (decays) into a simpler and more stable configuration. This decay results in the emission of particles and electromagnetic energy.

The "half-life" of a radioactive element is defined as the time required for half of any given amount of the element to decay. Half-lives range from more than a billion years for some nuclei to less than one billionth of a second.

Chapter 15

THE QUANTUM LEAP

"*A new scientific truth does not triumph by*
convincing its opponents and making them
see the light, but rather because its opponents
eventually die, and a new generation grows up
that is familiar with it."

Max Planck, German physicist (1858-1947)

SPECTRAL LINES

Around 1666, you will recall, Newton discovered that, after a narrow beam of sunlight passes through a glass prism, a series of colored bands appears on a screen in the familiar sequence of the rainbow; he called it the "spectrum" of sunlight. In 1802, another Englishman, W.H. Wollaston, observed that there were some dark lines among the intense colors of this spectrum.

In 1814, a German physicist, Joseph von Fraunhofer, modified Newton's experiment by replacing the screen with a telescope through which he could view sunlight that had passed through a narrow slit and a glass prism. Through his "spectrometer", he could see a multicolored strip displaying here and there "an almost countless number of strong or weak lines which are darker than the rest of the colored image; some appeared to be almost perfectly black."

An explanation for the Fraunhofer lines was found by another German physicist, Gustav Robert Kirchhoff. In 1859, he did an experiment in which light from the Sun was made to pass through a flame heavily charged with salt (sodium chloride), before going through the narrow slit of a spectrometer. Within the solar spectrum, two bright lines appeared in place of the two closely spaced dark lines that Fraunhofer had labeled the "D lines".

Kirchhoff concluded that sodium present in the Sun's atmosphere absorbs the radiation corresponding to the two D lines. In his experiment, the sodium in the flame had emitted radiation that restored the same two lines. The other Fraunhofer dark lines could be similarly attributed to the absorption by other elements present in the Sun's atmosphere.

Shortly after, Kirchhoff announced the two fundamental laws of the new science of "spectroscopy":

- Each chemical element has a unique spectrum, which consists of a number of distinctive lines. **Figure 15.1** shows, for instance, the spectral lines that are unique to sodium, hydrogen and mercury. Each line corresponds to a particular frequency of electromagnetic radiation.
- Each element is capable of absorbing the same radiation it is capable of emitting.

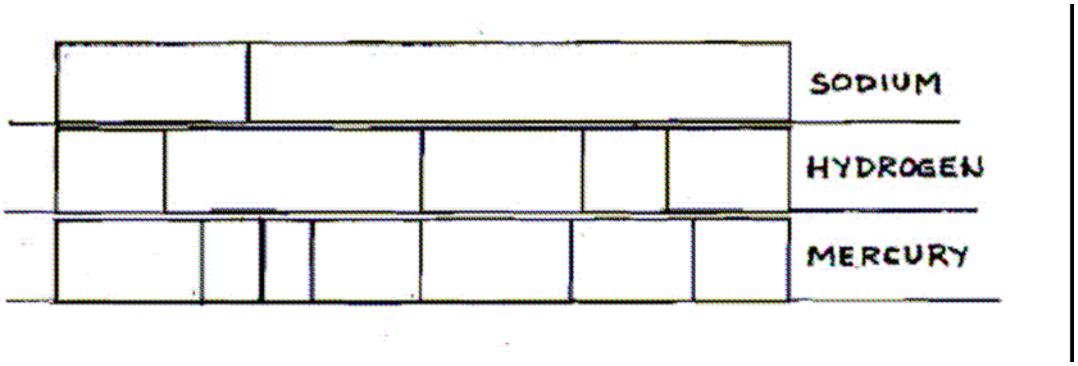

Figure 15.1 - Spectral lines of three elements

(From Cooper Plate 34.3 pg 433)

It was soon discovered that some of the lines unique to an element were in the infrared or ultraviolet, rather than the visible, portion of the spectrum.

At the time, physics had no explanation for the amazing fact that each element has a unique set of fingerprints. The answer came half a century later from quantum theory.

PLANCK'S QUANTUM

As we heat an object, say, the tip of an iron rod in a blacksmith's forge, the light it emits turns from dull red to bright red and then to white. If heated to a sufficiently high temperature, it turns bluish. What causes this gradual change of colors?

In the late 1800's, physicists studied extensively how electromagnetic radiation, including light, is emitted by a heated oven through a small hole. Since the molecules of the heated material inside the oven are vibrating, charged particles within the molecules are in accelerated motion, causing the emission of electromagnetic waves. These are repeatedly bounced around by the walls of the oven, until they eventually escape through the small hole As the oven is heated to higher temperatures, the hole will glow first dull red, then bright red, then white, then bluish, which is also what happens to the heated tip of an iron rod.

The radiation escaping from the hole can be analyzed experimentally to determine how the brightness of the light changes at various frequencies for a given temperature of the oven. Plotting the data results in curves like the two illustrated in **Figure 15.2**.

Attempts were made to predict mathematically the shape of these curves. Theoretical studies, however, concluded that, at any temperature, the intensity of the radiation emitted should continue to increase with the frequency, instead of decreasing after reaching a peak. Most of the radiation should be in the ultraviolet region of the spectrum and beyond. This result, first obtained in 1900, was clearly contrary to the experimental evidence. It became known as the "ultraviolet catastrophe".

In Germany, Max Planck (1858-1947) was able to derive a mathematical equation that was in agreement with the experimental curves. To succeed, however, he had to make a peculiar assumption, just tailored to get the desired result.

At the time, energy was thought to be continuous, i.e., it could increase or decrease by infinitesimal amounts. In terms of currency, we might compare it to very fine gold powder, with which we can pay in whatever amounts may be needed [5]. To get the desired result, however, Planck had to assume that energy is discontinuous, more like a currency with coins and bills of only certain denominations. To pay for some transactions, then, we have to use whole multiples of a single coin or bill of the right denomination.

Planck had to assume that heat converts into light not in any arbitrarily small increments, but in a granular fashion. The smallest quantity, or "quantum", of heat energy that can be converted into light of frequency f had to be equal to the frequency f times some constant h, h x f. (The constant h, called Planck's constant, would become as important in physics as the constant c, the speed of light in a vacuum.)

The total amount of heat energy converted into light of some frequency is a whole multiple of individual "granules" or quanta whose size is proportional to the frequency.

---

[5]. For gold powder to be truly continuous, its grains would have to be infinitesimally small.

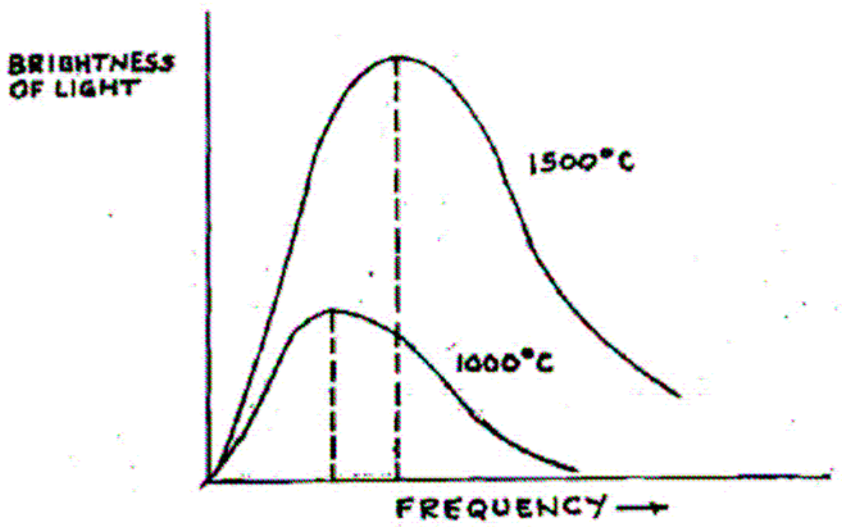

Figure 15.2 - Electromagnetic radiation emitted at two different temperatures

In the lower curve, the brightness of the emitted light first increases with frequency toward a peak, and then starts decreasing.

The upper curve describes what happens for a higher temperature of the oven. The range of frequencies becomes wider, the values of brightness are higher, and so is the frequency at which maximum brightness occurs. As the temperature rises, the curve shifts toward the right (in the direction from red to blue to violet). (from Poet pg 189)

We can see now, roughly, how Planck's assumption can explain the shape of the previous plots. After reaching a peak, each curve starts to decline because of the increasing investment of energy required to create a quantum at higher frequencies. Neither small coins nor very large bills are likely to contribute to the bulk of the money to be found in a cash register.

The assumption of granules or "quanta" of energy gave the desired result, but appeared to be totally unnatural to the physicists of the time, Planck included. In classical theory, the energy of a wave depended on the amplitude of the wave, not on its frequency. For years, the idea of the quantum was viewed - even by Planck - as a convenient "working hypothesis", a mathematical trick with no physical basis.

Planck, who died at 89 in 1947, endured many personal tragedies. Both of his daughters died in childbirth shortly after getting married. His elder son was killed in battle during World War I. His other son was implicated in an attempt to assassinate Hitler, and died a horrible death at the hands of the Gestapo. His house was destroyed by an air raid.

EINSTEIN'S PHOTON

The next application of Planck's quantum hypothesis was made by Einstein. He used it to explain a relatively obscure phenomenon called the "photoelectric effect".

In the late 1800's, it was known that, under certain conditions, when light was shined on a metal surface, it would cause electrons to be knocked off the surface. Some qualitative facts were known, but without any explanation:

- The photoelectric effect is easily produced with blue or ultraviolet light, but not with red light, which has a lower frequency.
- When electrons are emitted from the metal surface, their kinetic energy does not exceed some maximum value. If we increase the *brightness* of the light, more electrons are emitted, but their maximum kinetic energy does not increase. It will increase, however, if we increase the *frequency* of the light.

Einstein gave an explanation based on Planck's hypothesis, which he viewed, not as a mathematical trick, but as a profound truth applicable to all aspects of physics.

In one of his famous papers of 1905, he suggested that light was not only emitted in Planck's little granules of energy, but it was also absorbed in such granules. This granularity represented a property of light itself, independent of how it was emitted or absorbed. Einstein proposed a theory in which light can be likened to a hail of energy granules with no mass. These energy granules were later named "photons", each carrying a quantum of energy proportional to the frequency of the light.

The theory provided a convincing explanation of the photoelectric effect. A single electron will either absorb a whole photon or none at all. Only if the frequency of the light is high enough, will the "kick" delivered by the photon be large enough to enable the stricken electron to break free from the metal, after colliding with atoms, and have some kinetic energy left over.

If the frequency of the light is even higher, the photon's "kick" will be stronger and the stricken electron will have more kinetic energy left over, after breaking free. Increasing the intensity of the light, instead of its frequency, makes more photons available, but does not increase the strength of the "kick" each photon can deliver.

The puzzle of the photoelectric effect had been solved, but only to raise new questions. How could Einstein's particle theory of light be reconciled with the overwhelming evidence that supported the wave theory of light? How could light be both waves *and* particles? This was the first instance of the paradoxical wave-particle duality that would become a fundamental aspect of 20th-century physics.

When Einstein was submitted in 1913 for membership in the Prussian Academy of Sciences, his sponsors felt obliged to make excuses for what seemed then to be the least respectable of his many achievements, the photoelectric theory. Three years later, however, the American physicist Robert Millikan experimentally confirmed Einstein's quantitative predictions. (Ironically, he had started with the intent of proving them wrong.)

In 1921, Einstein received the Nobel Prize in physics for his "photoelectric law and his work in the field of theoretical physics". No mention was made of relativity, still a controversial subject.

BOHR'S ATOM

After Rutherford proposed an atom with a central nucleus, a natural next step was to imagine the atom as a miniature solar system. Instead of planets orbiting the Sun under the pull of gravity, one would have negative electrons orbiting the positive nucleus under the pull of electrical attraction.

The idea of such a parallel between the very small and the very large was very appealing, but totally unworkable. According to Maxwell's theory, since an electron would be continuously accelerated as it revolved around the nucleus, it should radiate light and thereby lose energy. This would cause it to spiral faster and faster toward the center; within billionths of a second, it would crash into the nucleus.

A new model of the atom was proposed in 1913 by a young Danish physicist, Niels Bohr (1885-1962). After receiving his doctorate at the University of Copenhagen, in 1912 Bohr went to work under Rutherford in Manchester. Here, within three months, he laid the foundations for a new theory of matter that would break away completely from what is now called classical physics.

Bohr, who was familiar with Planck's work, started with the simplest of all atoms, the hydrogen atom, which consists of a single proton and a single electron. To stay consistent with experimental facts, he combined some well-established principles of classical physics with some nonclassical hypotheses, which, like Planck's, were purposely tailored to fit the situation. The theory that resulted was a hodgepodge, but it worked.

For his model of the hydrogen atom, Bohr made three assumptions:

- Electrons move in circular orbits, but only orbits with certain radii are allowed.
- As long as an electron is moving in an allowed orbit, it does not radiate energy, as it should according to Maxwell's theory.
- Energy is radiated only when an electron leaps from one allowed orbit to another.

For the smallest allowed orbit of the electron, orbit #1, Bohr computed its radius to be about 5 billionths of a centimeter, about 50,000 times the radius of the nucleus. When the electron is in this orbit, it is said to be in its "ground state". It is as close as it can ever get to the nucleus.

For orbit #2, the radius is 4 (2x2) times the minimum radius; for orbit #3, it is 9 (3x3) times; for orbit #4, it is 16 (4x4) times, and so on.

In Bohr's theory, there is an "energy level" associated with each allowed "state" or orbit of the electron. You can think of this energy level as what the electron "owes" to become free from the nucleus. It is considered to be negative energy in the same sense that debt can be viewed as negative wealth: the smaller your debt, the larger your "wealth".

When the electron is in orbit #1, it is at its lowest energy level (highest debt). In orbit #2, its debt is 4 (2x2) times smaller; in orbit #3, it is 9 (3x3) times smaller, and so on. The larger the radius of the orbit, the more loosely bound the electron is.

The electron is normally in its smallest orbit at its lowest energy level. Let us assume that, while the electron is in this state, the hydrogen atom is disturbed somehow, for instance, by collisions with other atoms, or by photons bombarding it. If the electron receives a shot of energy equal to the difference between energy levels #1 and #2, it will be "excited" to jump to level #2.

Conversely, when the electron makes a transition from level #2 back to level #1, it releases a photon whose energy is the difference of the two energy levels (and whose frequency is its energy divided by Planck's constant).

More generally, if the electron receives a shot of energy in just the right amount, it can make a "quantum leap" from whatever energy level it already occupies to a higher one, not necessarily consecutive. Later, it can slide down to some lower level, not necessarily consecutive, and emit a photon, whose energy is the difference of the two levels.

We might say that, in making its transitions up or down, the electron, instead of a continuous ramp, has available a ladder with discrete rungs at uneven distances. When it leaps from one rung, it must land on another.

In Bohr's theory, consecutive orbits of the electron are assigned consecutive integer values of what is called the "quantum number n". That is why we have spoken of orbits #1, #2, #3 and so on. For each value of the "quantum number n", Bohr was able to compute the values of the associated radius and energy level. **Figure 15.3** shows the "rungs" of the energy “ladder" of the hydrogen atom for values of the quantum number n from 1 to 5.

Bohr's theory could explain with remarkable accuracy many details of the hydrogen spectrum, including spectral lines that were discovered outside the visible spectrum.

In 1913, Bohr published his theory of the hydrogen atom. In 1916, at 31, he returned to Denmark to accept a professorship at the University of Copenhagen. In 1920, he became director of the newly founded Institute of Theoretical Physics, which over the years attracted many of the world's leading physicists to Copenhagen. In 1922, he was awarded the Nobel Prize.

When Denmark was invaded by Germany in 1940, Bohr refused to cooperate with the Nazis. In 1943, under threat of imminent arrest, he was whisked away with his family to Sweden on a fishing boat in the middle of the night by the Danish underground. Eventually, he made his way to the United States, where he worked at the Los Alamos Laboratory on the Manhattan Project, which built the first atom bomb. He later returned to Copenhagen, where he died in 1962.

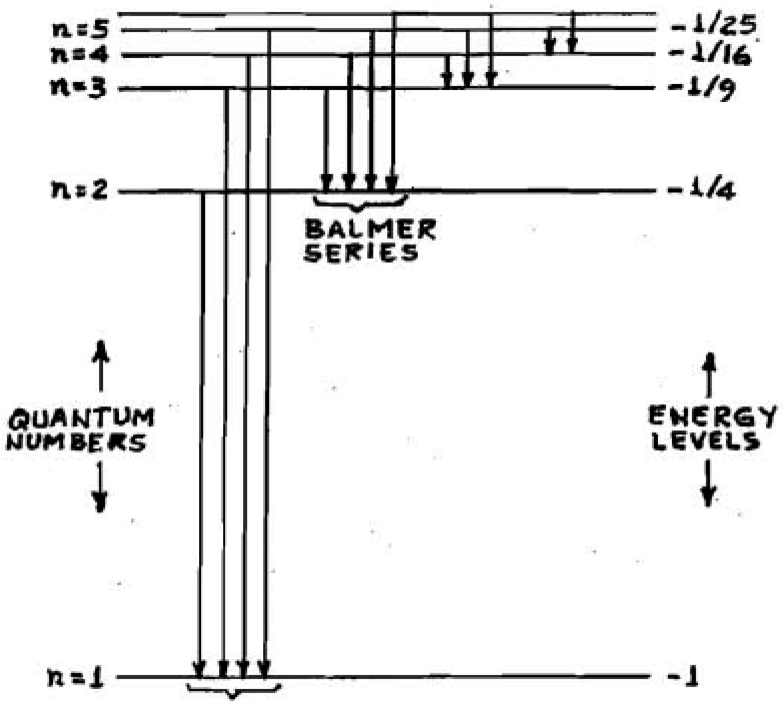

Figure 15.3 - Energy level diagram for hydrogen in Bohr's model

For the various values of the quantum number n, the corresponding energy levels are shown relative to the lowest level, to which is accordingly assigned the value -1.

The arrows represent downward transitions between energy levels. Each corresponds to a line in the hydrogen spectrum. The four transitions down to level #2 correspond to the four spectral lines shown in Figure 15.1. These are the only ones to fall within the visible spectrum.

(From Hey Figure 4.13 pg 48)

Chapter 16

QUANTUM MECHANICS

"Those who are not shocked when they
first come across quantum theory
cannot possibly have understood it."

Niels Bohr

"I think I can safely say that nobody
understands quantum mechanics."

Richard Feynman

Empirical Refinements of Bohr's model

Bohr's achievement in 1913 was remarkable, but his theory was by no means complete. It did not account, for instance, for the fact that some spectral lines are brighter than others; nor could it be applied to more complex atoms.

Bohr tried for many years to develop a mathematical theory for helium, which has the second simplest atom (a nucleus and two electrons). When more than one electron is involved, the situation becomes much more complicated because, in addition to the attractive forces between the positive nucleus and the electrons, one must consider the repulsive forces among the negative electrons.

In the years that followed Bohr's discovery, physicists extended his theory, in order to account for various features of atomic spectra. Defining the "state" of an electron became more complicated than specifying one of a number of concentric circular orbits, all on the same plane.

Whereas a single "quantum number" had sufficed for Bohr's theory, four quantum numbers, eventually, had to be used. Three were associated with the size, shape and orientation of the allowed electronic orbits, which could be circular or elliptical. The fourth quantum number was associated with a spinning motion of the electron about its axis.

Quantum numbers were assigned to the electron using rules that were discovered empirically. What was lacking was a sound mathematical foundation.

Einstein's Statistics of Electron Excitation

In 1916, after he completed his General Theory of Relativity, Einstein turned his attention again to quantum theory. He used statistical techniques and Bohr's model of the atom to study the behavior of huge numbers of atoms. According to Bohr, in a hot gas, as countless atoms constantly collide with one another, electrons are excited to higher energy levels; later, they fall back to lower levels and emit radiation (photons).

Einstein's statistical approach could explain why some spectral lines are brighter than others, the reason being that some transitions between energy states are more likely to happen than others.

After Maxwell, reality seemed to consist of empty space populated by two totally different kinds of "things": particles and waves. Particles were thought of as being point-like and having such properties as energy, momentum, mass and, possibly, an electrical charge. Waves, on the other hand, were thought of as being spread out in space, and having such properties as energy, amplitude, frequency, wavelength and speed of propagation.

In 1905, Einstein upset this sharp distinction. As we saw, he proposed that light - whose wave-like nature had been accepted after Young's 2-slit experiment - could be emitted or absorbed only as discrete quanta of energy. In 1909, he started talking of "point-like" quanta or massless particles, later called photons.

Einstein's statistical calculations of how matter absorbs or emits radiation explained how momentum could be transferred from a photon to an electron. It was necessary, however, to assume that each photon carried with it a momentum which was proportional to frequency.

Momentum had been considered before to be a particle-like property, normally expressed as mass x velocity. The massless photon was seen now as something that had both a particle-like property (momentum) and a wave-like property (frequency).

Prince de Broglie

In 1924, while still a graduate student in physics, a French nobleman, Louis de Broglie (pronounced de Broy), proposed a brilliantly bold conjecture. If light waves can behave like particles, why shouldn't particles, such as electrons, behave like waves? Why couldn't an electron be a wave?

Prince de Broglie was born in 1892 in one of Europe's most aristocratic families. For centuries, his family had contributed diplomats, cabinet ministers and generals. After he entered the University of Paris in 1910, inspired by his brother, who was a physicist, he developed an interest in physics and in Einstein's work. His idea of a wave/particle duality became the basis of the doctoral thesis he submitted in 1924.

De Broglie proposed that electrons as well as photons have associated with them both a momentum and a wavelength[6]. He could not say what was the physical significance of an electron's wavelength, but he could show an interesting connection between it and Bohr's orbits. The circular orbits that were allowed in Bohr's theory turned out to be those that could contain exactly a whole number of wavelengths.

Even for a doctoral thesis, de Broglie's proposal was too original to be comfortably accepted by the examiners without any experimental support. A copy of the thesis was sent to Einstein, who commented "It may look crazy, but it really is sound". (Years later, shortly before his death, commenting about some controversial new theory, Bohr remarked that the theory was certainly crazy, but wondered whether it was crazy enough to have merit.)

In 1927, confirmation of de Broglie's electron wave was provided by experiments performed by Clinton Davisson at Bell Telephone Laboratories in New York. In 1929, de Broglie became the first to receive a Nobel Prize for a doctoral thesis.

In 1937, Davisson shared the Nobel Prize with George Thomson, who independently had found confirmation for the electron wave. Interestingly, in 1906, J.J. Thomson had received the Nobel Prize for proving that electrons are particles; 31 years later, his son received the Nobel Prize for proving that electrons are waves! The reader may be totally mystified by the idea of a "particle" behaving like a "wave", but so was de Broglie who proposed the idea, and so were the physicists of his day.

QUANTUM MECHANICS

At the beginning of 1925, the quantum theory that started with Bohr in 1913 was still a hodgepodge of hypotheses and cookbook rules. In the 12 months following June 1925, quantum theory finally gained a firm mathematical footing.

Not one but two theories emerged, independently of one another. Although their approaches seemed at first sharply different, the two theories were later proved to be equivalent aspects of a single theory. Now called "quantum mechanics", this combined theory was the creation of a new generation of physicists, mostly born since Planck's discovery of the quantum. Their youth made them scientific revolutionaries willing to break away from classical physics. Bohr was the undisputed guiding spirit of this extraordinary development, and Copenhagen became its center.

The Gottingen Trio

The first of the two mathematical theories was developed by three German

---

[6] For light or, more generally, electromagnetic waves, we can talk in terms of either frequency or wavelength because the product wavelength x frequency is a constant, the speed of light c.

physicists at the University of Gottingen: Werner Heisenberg (1901-1976), Max Born (1882-1970), and Pascual Jordan (1902-1980).

Heisenberg, who was at the time a research assistant to Max Born, proposed a radical reinterpretation of the basic concepts of mechanics with regard to atomic particles. The new approach was guided by the principle that a physical theory is obligated to consider only those things that can actually be observed. There is no way of observing directly an electron orbiting around the nucleus of an atom. What can be observed are spectral lines, which are interpreted in terms of electrons jumping from one energy level to another.

The notion of tiny balls in orbits is a convenient mental image that is superimposed on the actual observations, because that is how we see things moving in our everyday world. In developing his theory, Heisenberg was willing to abandon convenient analogies and mental images, and to follow a totally abstract approach.

The breakthrough came in June of 1925 while he was recovering from a severe attack of hay fever on a North Sea island. There, away from distractions, he was able to concentrate on the new ideas that were forming in his mind. Within three months, the collaboration of Heisenberg, Born and Jordan resulted in the publication of a comprehensive paper.

Schrodinger's Equation

Only months later, Erwin Schrodinger (Austrian, 1887-1961) published a different mathematical theory he had developed independently of the Gottingen group. He had been inspired by Einstein's support of de Broglie's theory. Unlike the Gottingen group, he was guided by mental images of atoms with electrons as waves.

Before long, the two theories were proved to be fully equivalent. Being much simpler to use, Schrodinger's equation became the preferred mathematical tool of quantum mechanics to solve atomic problems. The equation gave a very good account of the spectrum of the hydrogen atom in a way that was mathematically consistent. But even for the very simple case of hydrogen, the mathematics involved is very difficult. To convey just a flavor of this complex theory, it will suffice to summarize the results as follows.

Each allowed "state" of the electron is some configuration identified by a particular set of values for three "quantum numbers" n, l (the letter l) and m. These are whole numbers related by simple rules, which stem directly from the equation.

For the hydrogen atom, Schrodinger obtained energy levels that are the same as Bohr's and depend only on the quantum number n. Except for n=1, however, more than one state is associated with the same energy level, see **Figure 16.1**.

For each state (i.e., for each set of values for the three quantum numbers n, l and m), instead of a circular orbit, Schrodinger's equation yields a so-called "orbital", a mathematical expression that is unique to that state. There is a single orbital for the lowest energy level (n=1). As Figure 16.1 shows, the number of orbitals increases with the energy level.

Usually called a "wave function," [7] an orbital describes a standing wave of some particular shape. What is the significance of this "wave"? In Schrodinger's view, the electron was not a particle but a wave. Instead of having its entire mass and charge concentrated at some point, the electron would have that same mass and charge smeared over some region of space. The density of mass or the density of charge at any point would be proportional to the square of the amplitude of the wave at that point.

This interpretation, however, is not consistent with the fact that, whenever an electron is observed, it always comes as a whole lump with a fixed charge and a fixed mass.

Born's Probability Interpretation

To Schrodinger's wave amplitude, Max Born gave a radically different interpretation, which has become one of the cornerstones of quantum mechanics.

He proposed that the square of the "wave function" at some point is a measure, not of a mass density or charge density, but of the *probability* of detecting the electron, if we were looking for it at that point. In this interpretation, nothing will be found at that point most of the time, but when something is detected there, it will be a whole electron.

Schrodinger's wave function does not tell us where the electron is at any given moment, only what is the probability of its being at various places, depending on its energy level. The electron could be anywhere: it is extremely likely to be in some places, and very unlikely to be in others.

The probability is highest in the immediate vicinity of the nucleus; it becomes inconsequentially small even at moderate distances. For instance, the probability of the electron being farther than about 5 hundredmillionths of a centimeter from the nucleus is 1 in 30,000. The more energy the electron has, the higher the probability that it will be found at greater distances from the nucleus. Each orbital is a particular pattern of how the probability of detecting the electron varies from point to point.

---

[7]. In mathematics, the term "function" is used to refer to a mathematical formula, which states how some variable quantity depends on one or more other variables, such as time or position.

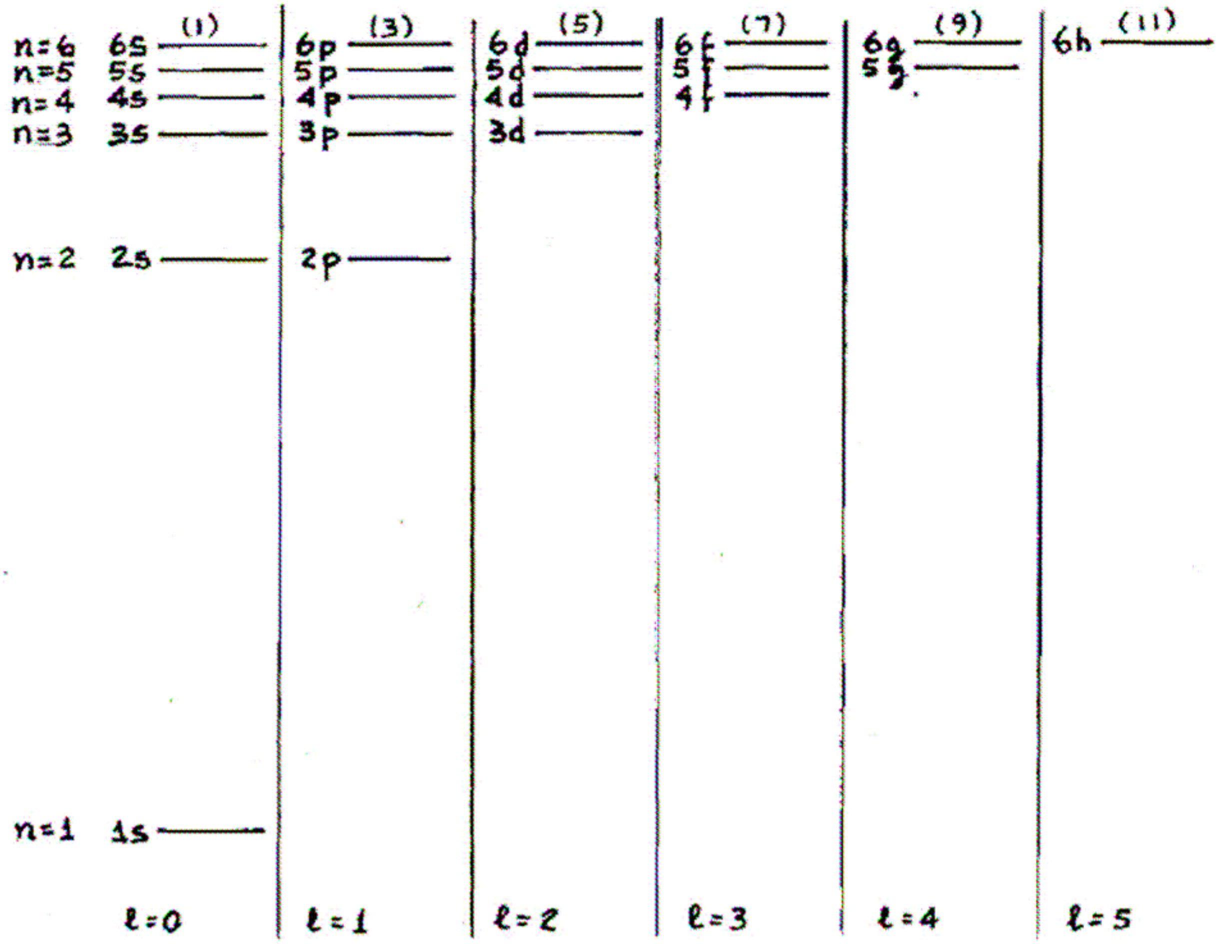

Figure 16.1 - Energy level diagram for hydrogen in Schrodinger's model

As shown above, for n=1, we have a single state possible, called the "1s" state.

For n=2, we have a single "2s" state, and three "2p" states; the latter are represented by the single line labeled "2p".

As a final example, for n=3, we have a single "3s" state, three "3p" states, and five "3d" states, for a total of 9 states.

Note that the interpretation of atomic behavior is still based on a ladder of energy levels, although the ladder is no longer as simple as the one we first saw for Bohr's model of the hydrogen atom. Electrons climbing up or down the ladder still account for the absorption or emission of radiation at discrete frequencies.

Schrodinger's equation can be applied not only to an electron bound within an atom, but, more generally, to a free electron, a photon, a proton, or any of the many “quantum entities” (particles) that have been discovered. A particular "wave function" is associated with a quantum entity. This mathematical expression describes a so-called "psi" wave, which may be a standing or traveling wave. It is not, however, an ordinary wave because it does not carry any energy. The square of its amplitude at some point in space represents the probability that the quantum entity will be detected, if we look for it there.

Except for not carrying energy, "psi" waves behave like ordinary waves. In particular, they can interfere with one another. When two or more waves are involved, they are first combined in the usual interference fashion, and then the square of the resulting amplitude is computed. The peculiar, but very important, result is that combining "psi" waves may result in diminished or even canceled probabilities.

The role of probability is the great divide that separates quantum mechanics from all preceding theories of physics. Newtonian mechanics was deterministic: given the position and velocity of a particle at some point in time, and given the forces acting on the particle, its path for all past and future times can be completely determined, at least in principle. The motion of the planets around the Sun provides a classical example.

Quantum mechanics, on the other hand, asserts that it is possible to predict only the probabilities of events. Instead of predicting a specific path for a particle, it gives a distribution of probabilities of where the particle might be.

Heisenberg's Uncertainty Principle

Another major blow to classical physics was delivered by Heisenberg in 1927 with his famous *Uncertainty Principle*, which he derived from the fundamental equations of quantum mechanics.

He was the first to point out that the laws of quantum mechanics imply a fundamental limitation to the accuracy of experimental measurements. This limitation has nothing to do with any flaw in the design of the experiment one might perform, or of the instruments one might use. It reflects a fundamental aspect of reality itself.

Heisenberg's uncertainty principle states that it is not possible to measure simultaneously the momentum (mass x velocity) and the position of a particle with whatever accuracy we might wish.

There is always some error or uncertainty associated with any measurement of position, and also some uncertainty associated with any measurement of momentum. What Heisenberg discovered is that these two uncertainties are unavoidably linked together. The more accurately we try to measure the position of an electron, the more uncertain becomes our knowledge of its momentum, and vice versa. In standard centimeter-gram-second units, Heisenberg's principle states that

(uncertainty in position) x (uncertainty in momentum)
must be larger than, or equal to about one billion-billion-billionth.

Note that, if one uncertainty is very small, the other must be large enough for the product of the two uncertainties to be larger than, or equal to, the very small number above.

Because of this very small value, the uncertainty principle affects only the submicroscopic world. When masses and distances have ordinary values, the quantum theory and the classical theory give results that are essentially identical. It is only with masses in the order of the electron mass, and distances in the order of atomic distances that we cannot neglect quantum effects.

Heisenberg illustrated the uncertainty principle by analyzing how we might go about to determine the position of an electron. To do so very accurately, it is necessary to use light with a very short wavelength. But a very short wavelength means a very high frequency, which in turn, by Planck's law, means a very large energy quantum.

Shining a light of very high frequency on a particle amounts to bombarding the particle with high-energy photons, which are going to deliver a very large kick, changing the particle's momentum by an indeterminate amount. Unavoidably, we have disrupted the very conditions we were trying to measure.

Conversely, if we want to know the momentum very accurately, we must use light that will deliver a very small kick, which means low frequency, long wavelength and, consequently, a large uncertainty in the measurement of position.

Dirac

In 1928, the most complete mathematical formulation of quantum mechanics was developed by the English theoretical physicist Paul Dirac (1902-1984), who shared in 1933 the Nobel Prize with Schrodinger.

Shy and withdrawn since childhood, Dirac became notoriously reserved as an adult. He seldom spoke and, when he did, chose his words with utmost care. He preferred to work alone. His favorite pastime was solitary walks. At the age of 30, he was appointed to the chair of mathematics at Cambridge, which had been occupied by Newton. His achievements are considered comparable to those of Newton, Maxwell and Einstein.

To arrive at his equation, Schrodinger had not taken into account Einstein's Special Theory of Relativity, which comes into play because of the very high speed of particles. Later, he and others tried to take relativity into account, but with meaningless results. Dirac succeeded using a very ingenious method. His mathematical solution showed that, in addition to the three quantum numbers n, l and m of Schrodinger's equation, an electron requires a fourth quantum number, which had been introduced earlier, but only on empirical grounds. This fourth quantum number is associated with a spinning motion of the electron about its axis, in either direction. It has only two values, which are called at times "spin up" and "spin down".

Quantum Mechanics and the Periodic Table

One of the many achievements of quantum mechanics was its ability to account, at least qualitatively, for many features of the periodic table of chemical elements. It can account, for instance, for how the number of elements varies from one row of the periodic table to another. (As shown at the end of Chapter 5, the numbers of elements in the 7 rows are: 2, 8, 8, 18, 18, 32 and 20). That these numbers differ has to do with how the number of energy states increases with the quantum number n.

The theory can also explain why elements in the same column of the periodic table have similar properties. Such elements have the same number of electrons outside an outermost "shell" in the atom, even though they have different numbers of electrons inside this "shell". It is the equal number of outlying electrons that is mainly responsible for the similarity of properties for elements in the same column.

With its ability to explain the chemical behavior of atoms, quantum mechanics achieved the unification of the previously distinct fields of physics and chemistry.

Chapter 17

QUANTUM INTERPRETATIONS

"*God does not play dice with the universe.*"

Einstein

There is no question that quantum mechanics is a theory that works. It has been enormously successful in explaining phenomena at the atomic and subatomic level. Semiconductors, transistors, microchips, lasers and nuclear power are among its fruitful applications. What the theory really means, however, has been, and still is, a mystery. Its concepts are difficult to understand as well as to believe. Often, they are in conflict with common-sense notions derived from what we observe in the everyday world.

Some forty years after the birth of quantum mechanics, Richard Feynman, American Nobel-Prize physicist and renowned teacher (1918-1988), remarked during a lecture:

"I think I can safely say that nobody understands quantum mechanics. So do not take the lecture too seriously, feeling that you really have to understand in terms of some model what I am going to describe, but just relax and enjoy it. I am going to tell you what nature behaves like. ... Do not keep saying to yourself, if you can possibly avoid it, 'But how can it be like that?' because you will get … into a blind alley from which nobody has yet escaped. Nobody knows how it can be like that." [1]

A number of different interpretations have been proposed for quantum mechanics. Although mutually exclusive, they all meet the requirements of an acceptable theory. All accurately account for known experiments; all correctly predict the outcome of new experiments.

The “orthodox” view of quantum mechanics has been the "Copenhagen Interpretation", so called because its main proponent, Bohr, worked in that city. It has been taught in colleges to future physicists.

The Copenhagen Interpretation ruled for more than 50 years until well into the 1980s. The great majority of physicists have given at least lip service to this interpretation. What has mattered all along to most physicists is that quantum mechanics does indeed work as a practical tool. In more recent years, however, increasing efforts have been made to find alternative interpretations. These efforts, however, have in no way diminished the mysteries of the quantum world.

THE COPENHAGEN INTERPRETATION

Bohr first presented the basic concepts of this interpretation in September of 1927. The whole theoretical package, which incorporated the views of Bohr, Born, Heisenberg and others, was essentially completed in the early 1930s.

According to Bohr, a fundamental quantum entity, such as an electron or a photon, is neither a particle nor a wave, but both wave and particle pictures are necessary to explain the quantum world. They are *complementary* aspects of the electron's complex nature. Depending on the nature of the experiment under consideration, an electron (or photon, or proton) can be observed to behave in some cases as a particle, in others as a wave, but *never as both*. This is Bohr's "principle of complementarity".

As we saw, Born interpreted Schrodinger's "wave function" as the basis for a probability wave. According to Bohr, an electron (or any other quantum entity) does not really exist in the form of a particle, when nobody is looking at it. It exists merely as a superposition of states. The probability wave, we might say, gives us the shape and density of a swarm of ghostly electrons distributed somehow in space. When an observation is made, only one of these ghostly electrons materializes. Suddenly and mysteriously, the probability wave "collapses" and the electron appears, as a whole particle, at one particular point.

The objective existence of an electron at some point in space, independent of actual observation, has no meaning. The electron seems to spring into existence as a real object only when we observe it.

Heisenberg's uncertainty principle tells us that it is not possible to measure at the same time the momentum and the position of a particle with whatever precision we might wish. The Copenhagen interpretation goes further. What is questioned is whether, prior to some observation, there is a particle that exists on its own and has a precise momentum and a precise position.

According to classical physics, the entire universe consists of nothing but ordinary objects, where by "ordinary object" we mean an entity that possesses attributes of its own, whether they are observed or not. Quantum mechanics, instead, denies the common-sense notion that ordinary objects are themselves made of ordinary objects. In Heisenberg's words, "Atoms are not things".

There are "static" attributes, such as mass or charge, which do intrinsically belong to an electron, and distinguish it from other kinds of particles. On the other hand, according to the Copenhagen Interpretation, there are "dynamic" attributes such as position or momentum, which seem to depend on how they are measured. A dynamic attribute seems to belong jointly to the electron and the measuring device. There is no hidden value of position that the electron "really" has when it is not being measured.

In classical physics, a system of interacting particles can be compared to some elaborate clock mechanism, which functions on its own, whether or not it is being observed. In quantum physics, the observer is viewed as interfering with the system under observation to such an extent that the system cannot be thought of as having an independent existence of its own. In the microworld, the very act of observing something changes it: the observer is very much part of the experiment.

The reality that is observed cannot be divorced from the observer and what he chooses to observe. In Bohr's view, there are no atoms, only measurements. The atom is a creation of the human mind to bring some order into the chaotic pattern of observations. Only what we observe can be considered real.

About any quantum experiment, according to Bohr, all that can be said is that, if physicists set up an experiment in a certain way and make certain measurements, they will get certain results. What are regarded as physical attributes of an electron are actually relationships between electrons and measuring devices. These properties "belong" to the whole experimental setup, not to the electrons.

To account for atomic phenomena, one must abandon the notion that the movement of a particle can be represented by a continuous succession of positions in space along a particular path. A subatomic particle does not appear to follow a well-defined trajectory at all. A particle seems to be able to go from one place to another without traversing the space in between!

The Copenhagen interpretation does not attempt to explain what might *really* be happening "behind the scenes" in the quantum world. It claims that *no* theory can explain subatomic phenomena in any more detail. It is not necessary to know how light can manifest itself both as particles and waves. It is enough to know that it does. Developing overall views about the nature of reality does not matter much. Pragmatically, what matters in physics is the development of mathematical equations that enable the physicists to predict and control the behavior of particles. In Bohr's words: "It is wrong to think that the task of physics is to find out how nature *is*. Physics concerns what we can *say* about nature." [2]

The Bohr-Einstein Debate

In October of 1927, some thirty eminent physicists met to discuss the meaning of the new quantum theory. Einstein, Planck, de Broglie and Schrodinger were among those who were disturbed by the direction the new theory was taking in the emerging Copenhagen interpretation.

To focus the issues, Einstein proposed a simple thought experiment in which we are asked to imagine a single electron passing through a very small hole. The probability wave associated with the electron is diffracted by the hole, and starts spreading in a hemispherical pattern toward a concentric hemispherical screen. There are two different viewpoints as to what actually happens.

In Einstein's view, the electron is a real particle with an existence of its own. Its associated probability wave must be supplemented by some still undiscovered process that will explain why the electron lands where it does. Otherwise, we would have to imagine some mysterious action-at-a-distance that would suddenly cause the wave to collapse. Instantly, the probability of the electron landing somewhere on the screen would become zero everywhere except at the point where the electron is detected. This instantaneous action-at-a-distance would violate the Special Theory of Relativity. Quantum mechanics, therefore, must be considered an incomplete theory.

In Bohr's view, instead, the quantum theory is complete. There is no additional process still to be discovered. We can talk only about what we can observe. There is no particle out there on its own; it materializes only when we look for it. There is no reality until that reality is perceived. Reality depends on what and how we choose to observe.

This was the beginning of an intense, but friendly, debate on the nature of reality between Bohr and Einstein, a debate that would continue for almost 30 years until Einstein's death in 1955. Over the years, Einstein proposed one thought experiment after another trying to undermine Bohr's interpretation. Again and again, Bohr was able to win the argument by showing how the thought experiment actually could not be carried as proposed by Einstein.

Ultimately, Einstein lost the debate. Reluctantly, he had to concede that Bohr's views were at least consistent. He refused to concede, however, that they were the last word. Someday, he was convinced, a new discovery would restore order to the world of particles.

At the heart of the debate was a fundamental concept of quantum mechanics: randomness. Quantum mechanics predicts only the probability of some result. Consider, for instance, the radioactive decay of a large number of identical atoms. The theory says that the decay is a completely random process. It can tell us precisely what percentage of the atoms will decay, but it cannot predict which particular atoms will.

Classical physics too acknowledged the randomness of many processes. For instance, it was accepted that, when a roulette wheel is spun, the ball will drop at random in one of the numbered compartments. In principle, however, it was believed that the winning number could be predicted if one knew exactly the location of the wheel at the instant the ball was dropped, the speed of the wheel at the time, and various other physical variables. The randomness of the game of roulette was seen as the result of our ignorance of certain "hidden variables".

The Copenhagen Interpretation, instead, asserts that there is something absolutely fundamental about randomness. Although one atomic nucleus decays and another does not, both were previously in identical states. There are no hidden variables that might explain why one nucleus decays and the other does not; just pure randomness. To Einstein and others, this violated a fundamental principle of science, the principle of causality, whereby every effect must have a cause. Einstein summarized his refusal to accept this with his famous remark "God does not play dice with the universe".

The idea of hidden variables was dealt a serious blow by John von Neumann (1903-1957), a Hungarian mathematician who moved to the United States in 1930. In 1932, he presented a mathematical proof, which stated that no theory based on hidden variables could ever properly describe the behavior of quantum entities. Von Neumann concluded that electrons cannot be ordinary objects, nor can they be made of ordinary objects that are presently unobservable. The existence of an ordinary reality underlying the quantum facts, he claimed, is mathematically incompatible with quantum theory. He was one of the greatest mathematicians of his day, and his conclusion went undisputed for more than 30 years

.

The EPR Paradox

In 1935 in Princeton, Einstein and two other physicists, Boris Podolsky and Nathan Rosen, proposed a new thought experiment, known as the EPR experiment from the initials of the three men. Their intent was to show that, following impeccable reasoning strictly in accordance with the rules of quantum mechanics, they were led to a conclusion so absurd that the theory had to be viewed as incomplete.

A simplified version of the EPR experiment was later proposed by the American physicist David Bohm. It starts with generating somehow a pair of protons that are in close proximity and "correlated" with one another in such a way that they have equal and opposite "spins". If one spin is "up", the other must be "down". The actual spins, however, remain indeterminate until some measurement is made.

Suppose now that the two protons move in opposite directions until they are far apart. When we decide to investigate one of them, we find that its spin is "up". Quantum mechanics requires that the other proton orient itself to acquire an equal and opposite "spin down", as if it knew instantly the spin status of its twin.

Here is the key point of this thought experiment. Even though the two protons may be now millions of miles apart, quantum mechanics tells us that the second particle must be affected by something we have decided to do far away to the first particle. Einstein and his collaborators believed that "no reasonable definition of reality could be expected to permit this".

What appeared to be violated was the common-sense principle of "local causes", or "locality" principle. What happens in some area cannot be affected by what an experimenter may decide to do in some other distant area, if the two areas are so far apart that there is not enough time for a light signal to connect the two events.

In Einstein's view, the proton that had moved far away was independently real and had all along some particular spin. He could not accept that this spin could be affected instantly by what was done far away to the other proton.

In Bohr's view, one could talk only about the spin that was measured when it was measured. Until a measurement is actually performed, the two protons must be regarded as a single totality, however far apart they may be.

The EPR experiment had brought out an unexpected implication of quantum mechanics, "non-locality". It became known as the EPR paradox, and was never resolved in Einstein's lifetime.

Bell's Inequality

In 1966, the Irish physicist John Stewart Bell (1928-1990) showed that von Neumann's proof, which denied the possibility of hidden-variable theories, was based on a false assumption. Bell proved that hidden-variable theories could be made to work, provided we accept non-locality.

A "local reality" - the kind of common-sense reality envisioned by Einstein - is defined as one that is "local" (no influence can propagate faster than the speed of light) ***and*** "real" (whether we observe them or not, real particles exist "out there" with well defined properties of their own).

This definition of "local reality" plays a key role in a thought experiment proposed by Bell. In principle, it could be carried out on many "correlated" pairs of photons emitted simultaneously from an atom in two different directions. Bell showed that, if we live in a "local reality", then some particular pattern A of measurements must occur more often than some other pattern B. This is the famous "Bell's inequality" (A occurs more often than B).

On the other hand, if this inequality is violated, then ours is not a "local reality", which means that influences can propagate faster than the speed of light and/or particles do not exist independently of our observations.

Bell's criterion to test for local reality was quite specific; the trick was to carry out the very difficult experiment called for. At the time, not even Bell thought his experiment was a practical possibility.

Over the next 20 years, however, a number of ingenious experiments were actually carried out along these lines. The most comprehensive and conclusive of these experiments were those done by the physicist Alain Aspect and his colleagues in Paris in the early 1980's. The results of the Aspect experiments (and others) show that the Universe is not both "local" and "real".

These experiments "tell us that particles that were once together in an interaction remain in some sense parts of a single system, which respond together to further interactions. Virtually everything we see and touch and feel is made up of collections of particles that have been involved in interactions with other particles right back through time, to the Big Bang, in which the universe as we know it came into being. ... The particles that make up my body once jostled in close proximity and interacted with the particles that now make up your body."[3]

FEYNMAN'S CENTRAL MYSTERY

One of Feynman's key contributions to quantum theory was the idea that, when going from place A to place B, a particle takes account of every possible route, however complicated it might be.

Feynman's version of quantum mechanics says that we must calculate the effects of all the possible paths from A to B and add them together. It is called the "sum-over-histories" approach. The resulting wave function is the same as the one derived from Schrodinger's equation.

In his famous lectures on physics, to illustrate the strangeness of the quantum world, Feynman chose the 2-slit experiment, the one Young had used in the 1800's to "prove" the wave nature of light.

Feynman described the 2-slit experiment as "a phenomenon which is impossible, absolutely impossible, to explain in any classical way, and which has in it the heart of quantum mechanics. In reality, it contains the only mystery. We cannot make the mystery go away by explaining how it works. We will just tell you how it works. In telling you how it works, we will have told you about the basic peculiarities of quantum mechanics." [4]

For both water waves and light waves, we previously saw how, after going through two suitably narrow slits (or small holes) in some barrier, a wave gives rise to two diffraction waves, which spread out as two patterns of semicircular ripples and interfere with one another. In the case of light waves, a pattern of alternating dark and light fringes can be generated on a screen or photographic film placed at some distance on the other side of the barrier.

Following Feynman, let us consider now another version of the experiment, using a barrier with 2 holes. Imagine, he said, a somewhat wobbly machine gun shooting a stream of bullets at some armor plate with two holes A and B, each somewhat larger than a bullet. As it fires, the machine gun sprays the bullets randomly in a cone-like pattern aimed at the two holes in the armor plate. On the other side, at some distance from the plate, there is a backstop, perhaps a thick wall of wood, on which are mounted, close together, many little boxes or bins that can collect the bullets that go through the holes.

Suppose now that, after covering hole A, we shoot bullets for, say, 30 minutes, and then count the bullets accumulated in each little box. We will find the bullets distributed in a certain pattern. The box directly opposite hole B, which is open, will contain the largest number of bullets; nearby boxes will contain fewer bullets, and boxes further out will contain even fewer.

We repeat now the experiment with only hole B covered. Again, we shoot bullets for 30 minutes, and then count bullets in the various boxes. We will find the same distribution pattern, except that it is centered now in front of hole A, which is open.

We repeat the experiment one more time with *both holes uncovered*, and then count what we find in the various boxes. As we would expect, the number of bullets in each box is the sum of what we found there in the two previous experiments with one or the other hole uncovered.

In his lectures, back in the early 1960s, Feynman described next what happens when electrons are used in a two-hole experiment. The electrons are emitted one by one, go through the two holes of a barrier, and land on a screen where they can be detected. At the time, this was only a thought experiment, and Feynman described how the electrons were expected to behave according to quantum mechanics. (Actual experiments later confirmed these expectations.)

If we perform the experiment with *either hole covered*, we get a situation similar to what we saw for bullets. This is what we would expect since we are inclined to think of electrons as very tiny bullets. Like the bullets, the electrons arrive as identical individual lumps, one by one, each landing at some spot that will be generally different from one instant to another. Their distribution pattern will be similar to that of the bullets.

The situation, however, is strikingly different when the experiment is performed with *both holes uncovered*. The electrons still arrive, one by one, as identical individual lumps, but their distribution pattern is not the sum of the patterns observed when either hole is covered - as it was for bullets. The pattern appears to be random. Slowly but surely, however, it keeps building up to a pattern of several dark and light fringes that is characteristic of interfering waves. There are spots on the screen that receive many electrons when only one hole is open, and yet receive very few when both holes are open!

With both holes uncovered, the distribution of electrons detected on the screen is not the sum of the distributions when only one hole is open. What we get, instead, is an interference of probability waves from the two holes.

We can easily understand that a wave, which is spread out, can pass through both holes, creating diffraction waves that interfere with one another. An electron, however, still seems to be a particle, even if it has also wave-like properties. It would seem natural to expect that an individual electron must go through either one hole or the other. And yet, as reflected by the interference pattern of dark and light fringes that gradually develops on the screen, an electron supposedly going through one hole behaves quite differently depending on whether the other hole is covered or not.

The "central mystery" of quantum mechanics is that the electron behaves as if it "knows" what the situation at the other hole is.

More surprises are in store, as we consider a variant of the original thought experiment. Suppose we leave both holes open, but we place a detector by each hole so that we can monitor whether an electron does go through that hole. With this arrangement, we always detect an electron going through one hole or the other, never both at the same time. But now the interference pattern disappears!

The act of observing the electron wave with a detector makes it collapse and behave like a particle when it is going through the hole. Actually we don't need two detectors. Even if we monitor only one of the holes, the interference pattern disappears. The electrons going through the second hole "know" that we are looking at the first hole and behave like particles.

The electrons not only "know" whether both holes are open, they also know whether we are "watching" them, and they behave accordingly! If we set up the experiment expecting to see particles that go through one hole or the other, the electrons behave like particles, very tiny bullets, and no interference occurs. If, on the other hand, we don't attempt to see which hole they go through, they behave like waves, and interference occurs. Quantum mechanics seems to show that the observer's consciousness affects the nature of the physical reality that is observed.

In the mid-1980s, a team working in Paris was able to generate single photons going through an actual 2-hole experiment, one at a time. On the other side of the holes, some distance away, a photographic plate records the arrival of each photon as a white dot. The pattern of white dots appears random at first. But, as more and more photons keep landing, one by one, the white dots start merging to form the typical pattern of wave interference: white stripes with dark stripes in between, just as predicted by quantum theory. Each photon seems to "know" where to land on the film to make its own contribution to the overall interference pattern!

In 1987, a Japanese team carried out an actual two-hole experiment using electrons. The results were exactly the same as for the experiment with photons. Similar results were obtained in the early 1990s by a German team using atoms of helium.

Delayed-Choice Experiments

In the late 1970s, the eminent American physicist John Wheeler proposed, as a thought experiment, a variation of the two-hole experiment using photons. Detectors that can monitor the passage of the photons are placed somewhere between the two holes of a barrier and the screen where they are detected. We can then see whether the photons are behaving like particles or like waves *after* they have gone through the holes, but before they land on the screen.

If we choose not to look at the photons as they pass and switch the detectors off, the photons will create an interference pattern on the screen. On the other hand, if we choose to "look" and switch the detectors on, quantum mechanics says that no interference pattern will be formed. Paradoxically, whether the photons behave as particles or as waves as they go through the two holes is determined *after* they have gone through! Furthermore, we can wait until after the photons have gone through the two holes, before we even decide whether to switch the detectors on or off. If we look, we don't get an interference pattern; if we don't look, we do.

An actual "delayed-choice" experiment along these lines was carried out in the mid-1980s by two teams independently, one American, the other German. It confirmed the predictions of quantum mechanics. "The behavior of the photons is affected by how we are going to look at them, even when we have not yet decided how we are going to look at them!" [5]

In the early 1980s, Wheeler proposed a cosmic variant of his thought experiment extending over a huge span of time. The experiment involves photons that would have a choice of two different routes to reach Earth from a distant star. They could go either way, or they could mysteriously split up and travel both ways at once. Which route they follow - starting out, say, a billion years ago - would depend on whether or not an astronomer on Earth presently decides to switch on a detector attached to his telescope.

According to Wheeler, our misconception about this thought experiment is the assumption that, in his words, "a photon had some physical form before the astronomer observed it. Either it was a wave or a particle. Either it went both ways ... or only one way. Actually quantum phenomena are neither waves nor particles but are intrinsically undefined until the moment they are measured. In a sense, the British philosopher Bishop Berkeley was right when he asserted two centuries ago 'To be is to be perceived'." [6]

Wheeler has gone so far as to suggest that the entire universe exists only because someone is watching it.

OTHER INTERPRETATIONS OF QUANTUM MECHANICS

In the mid-1980's, eight top quantum physicists were interviewed in a British broadcast program. They all had different views of quantum mechanics, each firmly convinced that his interpretation was correct and the others were impossible.

Only two additional interpretations, which differ substantially from the Copenhagen interpretation, will be mentioned. Both have lately begun to gain wider recognition.

Wholeness and the Implicate Order

At the end of the 1940s, David Bohm (American, 1917-1992) refused to comply, when asked to testify before the Un-American Activities Committee of the House of Representatives about the political views of some of his colleagues in the Manhattan Project. Two years later, he was tried for contempt of Congress. Although acquitted, he found it impossible to get a job in the United States. He settled at Birkbeck College in London, England, where he developed his quantum interpretation over the next four decades.

Bohm was not deterred by von Neumann's "proof", which denied the possibility of "hidden variables" theories. As de Broglie had attempted back in 1925, Bohm pursued a quantum theory based on hidden variables and real particles that have at all times an intrinsic position and velocity. Any attempt to measure these properties, however, will destroy information about them by altering a "pilot wave" associated with the particles. This pilot wave is aware of conditions existing everywhere in the Universe, and guides its particles accordingly.

Through the pilot wave, everything in Bohm's reality is connected to everything else, and is instantly affected by whatever happens to everything else. Bohm writes about a universe of "undivided wholeness" and distinguishes between an "explicate order" (the one the laws of physics refer to) and an underlying "implicate order".

"Many-Worlds" Interpretation

A different interpretation was developed by Hugh Everett in 1957, when he was a student working under the supervision of John Wheeler. His basic idea is that, whenever the universe is faced with a quantum choice, the entire universe splits into as many copies of itself as there are possible options. For instance, in the two-hole experiment, when an electron is faced with the choice of two holes, the Universe splits into two copies: in one universe, the electron goes one way, while in the other universe, the electron goes the other way.

This theory "requires an infinite number of universes, each splitting into infinitely more versions of reality every split second, as all the atoms and particles in the universe(s) are faced with quantum choices, and follow every possible route into the future at once." [7]

The many-worlds interpretation makes exactly the same predictions as the Copenhagen interpretation. It has, for some, the advantage of avoiding the most vexing questions of the Copenhagen interpretation: When does the collapse of the wave function occur? Is consciousness an essential factor in the collapse of wave functions? The collapse of the wave function makes only one option real. In the many-worlds interpretation, instead, each option becomes real but in a different universe.

Wheeler initially endorsed the idea. A few years later, he changed his mind because he felt it carried too much "metaphysical baggage". After some 30 years of obscurity, the theory has met new interest on the part of some cosmologists.

Strictly speaking, the Copenhagen interpretation requires the existence of an observer *outside* the whole Universe to collapse all the wave functions and make reality real. To avoid this, some cosmologists prefer to believe that there really are countless universes.

The variety of conflicting interpretations of quantum mechanics is a vivid illustration of how subjective even the judgments of scientists can be. All the theories we have discussed are bizarre, whether we talk of particles that are not real until they are observed (Bohr), or of particles each traveling by all possible routes at once (Feynman), or of pilot waves that can sniff everywhere in the universe (Bohm), or of ever splitting universes (Everett). What is "metaphysical baggage" to one physicist is valid theory to another.

Chapter 18

FUNDAMENTAL PARTICLES AND FORCES

Since the late 1920's, science has made great strides toward an amazingly comprehensive view of reality with new theories and discoveries about:

- The fundamental particles that make up *all forms of matter*, and the forces that act upon them.
- The origin, evolution and structure of *the universe as a whole*.
- The amazing organization of DNA, the substance within each living cell that encodes genetic information for *all forms of life*.

The discovery of DNA, which is beyond the scope of this book, is the greatest triumph of biophysics, the study of biological phenomena using the principles and techniques of physics.

This chapter describes the "Standard Model", the theory of physics that summarizes the current scientific understanding of elementary particles, the building blocks of matter, and of the fundamental forces by which they interact. The following chapter gives an overview of the universe as a whole.

PARTICLE-ANTIPARTICLE PAIRS

When Dirac combined special relativity and quantum mechanics in 1928, there was a feature of his mathematical solution that appeared meaningless at first. Later, however, it turned into a triumph of theoretical physics.

Dirac's basic equation, which describes the allowed energies for electrons, has two sets of solutions, one positive and one negative. The positive solutions appeared to describe normal electrons. The negative solutions, however, seemed to be physically meaningless. Instead of disregarding them, however, Dirac interpreted them as describing a particle like an electron, having the same mass and the same, but positive, charge. Such a particle, called a "positron", was discovered in 1932 by an American physicist, Carl Anderson, who was not aware of Dirac's theory.

Anderson made his discovery while studying cosmic rays. These are high-energy subatomic particles, which rain down on the earth's atmosphere from outer space. Colliding with atomic nuclei in the atmosphere, cosmic rays generate showers of particles that cascade toward the surface of the earth. In these showers, the enormous energy of the incoming rays is converted into matter, in accordance with Einstein's famous equation $E = m \times c^2$. Among the particles created are pairs of electrons and positrons.

Positrons are not found in ordinary matter, and are very short-lived. When an electron and a positron come together, they annihilate one another, and their combined mass is converted into a high-energy photon. Each of the two particles is said to be the *antiparticle* of the other.

Positrons were the first example of antiparticles to be predicted and later discovered. Except for photons, every subatomic particle detected so far is known to have an antiparticle. In each particle-antiparticle pair, the mass is the same for the two particles, but the electrical charge, or some other property, has opposite values. Photons, which are massless, have no antiparticle counterparts.

SUBATOMIC PARTICLES

In addition to the particles that make up the atom (electrons, protons and neutrons), studies have revealed the existence of hundreds of particles (plus their antiparticles) which are smaller than atoms. These particles do not appear under the low-energy conditions of everyday experience. Also, they quickly decay to the more familiar particles after tiny fractions of seconds.

All these particles can be understood in terms of two types of elementary particles: the "leptons" and the usually more massive "quarks", plus their antiparticles. Both types have no discernible structure; that is, it does not appear that they can be separated into smaller components. Both types appear to behave as points in space, with no measurable size.

Leptons

There are believed to be only six types of leptons (plus their antiparticles). Only two are involved with ordinary matter: the familiar electron, and the chargeless "neutrino." Their antiparticles are the positron and the antineutrino. (Neutrinos and antineutrinos have opposite "spins".)

Neutrinos were thought to be possibly massless. More recent experiments, however, suggest that they do have some mass, but very small.

Neutrinos and anti-neutrinos do not exist within atoms as electrons do. They play an important role in certain types of radioactive decay. In one type, for instance, a neutron changes into a proton, and emits both an electron and an anti-neutrino.

Quarks

There are believed to be only six types of quarks (plus their antiparticles). Only two types of quarks are involved with ordinary matter: the so called "up" and "down" quarks. The up quark has a charge of +2/3e, whereas the down quark has a charge of -1/3e.

Two up quarks and one down quark combine to form a proton with a net charge of +e. Two down quarks and one up quark, instead, combine to form a neutron with zero net charge.

Each of the six quarks comes in three types or "colors" (the term "color" is not to be taken in any literal sense).

Particle Families

The six quarks and six leptons are grouped in three families, each with two quarks and two leptons, as summarized by the table below, which omits all the associated antiparticles.

Table I

	Fermions (24) (each has an antiparticle for a total of 48)			
Family	Quarks (6) (each comes in three "colors" for a total of 18)		Leptons (6)	
#1	up quark	down quark	electron	(electron) neutrino
#2	charm quark	strange quark	muon	muon neutrino
#3	top quark	bottom quark	tau	tau neutrino

The first family consists of the four particles involved with ordinary matter. The second family consists of heavier versions of their counterparts in the first family. Similarly, the third family consists of heavier versions of their counterparts in the second family.

Hadrons

Unlike leptons, quarks are never found as isolated particles. Together with other quarks or antiquarks, they form particles called "hadrons", which divide in two groups called "baryons" and "mesons". Baryons are particles built from three quarks, such as the proton and the neutron. Mesons, on the other hand, are particles built from one quark and one antiquark.

Among the hadrons, only the proton appears to be stable: its lifetime is at least 10^{32} years[8]. A single neutron outside a nucleus, on the average, would

---

[8] In this and the following chapter, we will deal with very very large, as well as very very small, numbers. It becomes then very convenient to use a notation that avoids long strings of zeros.

A notation like 10^{19}, for instance, is used as a shorthand to represent 1 followed by 19 zeros

10,000,000,000,000,000,000

10^6 then represents one million (1,000,000); 10^9, one billion (one thousand millions); 10^{12}, one trillion (one thousand billions, or one million millions).

On the other hand, for very small numbers, as a shorthand, instead of writing, for instance,

0.0000000000000000001

last only 15 minutes before decaying. Inside a nucleus, however, the neutron's lifetime is much longer so that many nuclei can be stable, thus making possible a large variety of chemical elements.

The other hadrons, which number in the hundreds (plus their antiparticles), are all very short-lived, with lifetimes ranging from 10^{-8} to 10^{-23} of a second.

FUNDAMENTAL FORCES

There are believed to be only four fundamental forces in the universe. Two are the already familiar forces of gravity and electromagnetism. The other two are the "strong nuclear force" and the "weak nuclear force", which are active only within the nucleus of the atom.

Of all the forces, the force of gravity is the weakest by far; it is also the most pervasive because its effects can be extremely long-ranging. The electromagnetic force, which affects only charged particles, is about a billion billion billion billion (10^{36}) times stronger than the gravitational force.

The strong nuclear force

The strong nuclear force is the strongest of the fundamental forces. It is typically 100 times stronger than the electromagnetic force, and 100,000 stronger than the weak nuclear force. It is ineffective beyond a range of 10^{-17} of a centimeter. Within the nucleus of the atom, however, it rules supreme.

Baryons (which consist of three quarks) are the particles that feel the strong nuclear force. Mesons (which consist of one quark and one antiquark) play a role in transmitting the strong force, as discussed below. It is the strong nuclear force that binds quarks within neutrons and protons, and holds protons and neutrons within the nucleus, overcoming the powerful repulsive forces among the positively charged protons. The strong nuclear force does not affect leptons.

The weak nuclear force

The weak nuclear force, which affects all particles, has a range shorter than that of the strong nuclear force by a factor of about 1000. It is responsible for certain types of radioactivity. A typical example occurs when a neutron transmutes into a proton.

Bosons

Up until the mid-1800s, before Faraday introduced the concept of field, a force was commonly believed to act at a distance. In the Standard Model, instead, the four fundamental forces are viewed as interactions between particles that

---

(1 preceded by 19 zeros, including the one before the decimal point), it is convenient to use the notation 10^{-19}

are mediated by the exchange of other particles called "bosons", which are different for different types of forces.

To get some general idea of how interactions between particles are mediated by an exchange of "messenger particles", think of two people skating toward a common point. One of the skaters carries a football; as he approaches the other skater, he throws him the ball. As a result, the two skaters will change direction, away from one another. One is deflected by the recoil from throwing the ball; the other, by the impact from catching the ball.

There are 12 types of bosons as follows:

- Eight kinds of "gluons", which mediate the strong nuclear force. These are all mesons.
- The familiar photon, which mediates the electromagnetic force.
- Three kinds of "weakons", which mediate the weak nuclear force. They are called W plus, W minus and Z zero, and are (electrically) positive, negative and neutral, respectively.

The 48 matter particles (fermions) previously listed and the 12 force particles above (bosons) constitute all the fundamental particles on which the Standard Model rests.

The gravitational force is believed to be mediated by yet another particle called the "graviton", but its existence has not been experimentally confirmed yet. Like the photon, the graviton would have no mass and no charge, and would travel at the speed of light.

The Standard Model, one of the great accomplishments of contemporary physics, was created by the contributions of many physicists. It is consistent with all experiments performed so far. It has, however, many shortcomings. It does not cover gravity. It does not explain why the various particles have masses that differ so widely. Its complexity leaves physicists striving for a simpler, more unified view of the world.

Chapter 19

A COSMIC PERSPECTIVE [9]

Modern cosmology, the study of the universe as a whole, has developed detailed theories about the origin, evolution, large-scale structure, and future of the cosmos. It owes its origins to Einstein's Theory of General Relativity, and to the discovery made in 1927 by the American astronomer Edwin P. Hubble (1889-1953) that the galaxies are receding from our Milky Way, and that the farther they are, the faster they are receding.

The fact that galaxies are moving away from our own galaxy in all directions suggested that, the further back in time we go, the closer together they must have been in the past, all the way back to some "primeval atom" containing the total mass of the universe. Cosmology accounts for the origin of the universe with the widely accepted "Big Bang" theory, according to which, about 10 to 20 billion years ago, the universe was born in a sudden outburst of energy of inconceivable intensity. Starting from a minuscule, immensely dense lump of energy-matter, the universe has been expanding ever since.

One would be easily tempted to imagine a gigantic explosion occurring at a particular place in a pre-existing empty space. The theory, however, tells us that the Big Bang gave birth not only to matter, but also to space and time. Galaxies today are receding from one another, not because they are rushing *through* space, but because the space between them is expanding. As an analogy, consider a balloon with many spots painted on it. As the balloon is inflated, all the spots are seen to move away from one another. From the viewpoint of any one spot, however, all the others appear to be receding from it.

A BRIEF HISTORY OF THE UNIVERSE

The First 100,000 Years (or so): The Age of Radiation [1]

Cosmologists believe that, from mathematical models, they can retrace the early history of the universe, but only (!) back to time $t = 10^{-43}$ of a second after the cosmic "explosion". At this moment in time, known as the "Planck time", the density of matter is an unimaginable 10^{94} grams per cubic inch, and the universe is only 10^{-33} of a centimeter across. The temperature has a staggering value of 10^{32} degrees K, ten trillion trillion times hotter than the core of an average star.

---

[9] More recent estimates may supersede some of the numbers used in this chapter. The grand panoramic view across time and space presented here, however, is likely to remain valid in its broad outlines.

As sketchily outlined below, the early history of the universe can be divided into a number of eras. Although their durations progressively increase from tiny fractions of a second to seconds to minutes to 100,000 years, each represents a major phase of cosmic evolution.

The GUT Era

This era lasted only about 10 billion-billion-billion-billionths (10^{-35}) of a second! The acronym GUT stands for "Grand Unification Theory". As proposed by a number of such theories, the three fundamental forces other than gravity were still unified (indistinguishable) during this era. Gravity is believed to have separated from the other three forces at the beginning of the GUT Era at time t = 10^{-43} of a second.

During this era, the universe is a "chaotic soup of energy-matter". In a fireball of radiation, pairs of matter and antimatter particles of various types spring into existence from extremely energetic photons, but are immediately annihilated in violent collisions that give birth to more particles and antiparticles.

The fundamental particles that make up the infant universe include photons, leptons, quarks, and their antiparticles.

The Inflation Era

The duration of this era is about 1000 billion-billion-billion-billionths (10^{-33}) of a second. As the universe expands, its temperature drops. When the cosmic temperature falls below a critical value of about 10^{27} degrees K, the universe undergoes a period of inflation (expansion) at a skyrocketing rate. By the end of this era, the volume of space has increased more than a trillion trillion times.

The strong nuclear force becomes differentiated from the electromagnetic and weak nuclear forces, which remain unified as the so-called "electroweak" force.

A break in the matter-antimatter symmetry results in about one billion and one particles of matter being produced for every billion particles of antimatter.

The Electroweak Era

The duration of this era is about one millionth of a second. The pull of gravity begins to slow down the expansion of the universe. The temperature drops to 10^{26} degrees K. The electroweak force divides into the electromagnetic force and the weak nuclear force. The separation of the four fundamental forces of the universe is now complete.

Leptons and antileptons evolve into variants such as electrons, positrons (or anti-electrons), neutrinos and anti-neutrinos.

The Quark Confinement Era

The duration of this era is about 2 seconds. The temperature is down to 10^{13} degrees K (more than a million times hotter than the core of the Sun).

Photons of very high energy are continually colliding with particles of various types, but they no longer have enough energy to prevent quarks from combining together to form protons and neutrons, the building blocks of future atomic nuclei. Photons, however, have still enough energy to smash any bonds between protons and electrons, thus preventing the formation of even the simplest atoms.

The cosmic fireball is dominated by photons, electrons, positrons, neutrinos and anti-neutrinos. Protons and neutrons represent a tiny minority. An anti-neutrino colliding with a proton produces a positron and a neutron. A neutrino colliding with a neutron produces an electron and a proton. Constantly bombarded, individual protons repeatedly change into neutrons and back to protons, while neutrons change into protons and back to neutrons.

Matter and antimatter continue to collide, mutually annihilating one another. Many of these events, however, produce photons that are no longer strong enough to create new particle-antiparticle pairs. Eventually, practically all antimatter disappears. In spite of the one-to-one annihilation of matter and antimatter particles, the slight excess of matter over antimatter produced during the Inflationary Era survives as the matter that constitutes the present universe.

The Neutrino Era

The duration of this era is 58 seconds. The temperature, which is now down to 10 billion (10^{10}) degrees K, continues to fall. By the time it drops to 3 billion degrees, no more electron-positron pairs are being generated, and the remaining ones are being annihilated. Positrons, like most other antimatter, gradually disappear. The antineutrino becomes the only antiparticle left.

No longer interacting with other particles of matter, neutrinos and antineutrinos form an almost undetectable "sea" that still fills the universe. Having no charge and nearly no mass, they can pass through matter as if it did not exist at all.

The Nucleosynthesis Era

The duration of this era is 4 minutes. The temperature is now 1,300 million degrees K; by the end of this era, it will be down to 600 million degrees. Photons can no longer prevent protons and neutrons from fusing into atomic nuclei under the pull of the strong nuclear force. They still have, however, enough energy to prevent the bonding of nuclei and electrons into atoms.

The most common combinations of protons and neutrons are the nuclei of hydrogen-2 (one proton and one neutron), helium-3 (two protons and one neutron) and helium-4 (two protons and two neutrons). The temperature of the expanding universe is dropping too quickly to allow more complex nuclei to form.

The Radiation Era

The duration of this era is about 100,000 years. At five minutes after the Big Bang, cosmic evolution slows down dramatically. The universe continues to expand and cool down. It is still too hot, however, for any stable atoms to form. As soon as a positive nucleus captures a negative electron, the electron is knocked out by an energetic photon. The universe is still dominated by radiation (photons).

When the temperature is down to 3,000 degrees K about 100,000 years after the Big Bang, photons have become weak enough that they can no longer prevent or disrupt the bonding of electrons with the simple nuclei of hydrogen and helium generated during the Nucleosynthesis Era.

At this point, the universe is a vast expanding cloud of gas consisting of approximately 75% hydrogen, 25% helium and traces of lithium. (As we will see, atoms of heavier elements will be formed only eons later in the blazing cores of stars.)

Practically all the electrons are now bound up in stable atoms. Since photons are no longer scattered by random encounters with electrons, they can finally burst free of matter. Space becomes transparent for the first time, and photons can travel unimpeded through the thin gas of the early universe.

From now on, photons and matter will rarely interact on a cosmic scale. As the universe continues to expand, the energy of photons continues to decline. Like the neutrinos before them, the primordial photons recede into a cosmic background. The universe has entered a new age no longer dominated by photons but by matter and gravity.

The Next 10-20 Billion Years: The Age of Matter

Galaxies

There is evidence that the cloud of hydrogen and helium that constituted the universe about 100,000 years after the Big Bang was extremely smooth and uniform in density. Yet, today the universe shows a very lumpy large-scale structure consisting of billions of stars grouped into "galaxies", which in turn form "clusters" grouped into "superclusters". How this large-scale structure of the universe came about is still an unsolved mystery.

It was only in the mid-1920s that most astronomers fully accepted the existence of galaxies beyond our Milky Way, which had been believed to constitute the entire universe. Before they could recognize the large-scale structure of the cosmos, astronomers had to develop ways of determining what was far and what was near, what was large and what was small - an exceedingly difficult task that was accomplished very gradually.

Galaxies typically contain from millions to hundreds of billions of stars. Their diameters range from 5,000 to more than 3 million light-years (a light-year is the distance traveled by light in one year, or about 6 million million miles).

The majority of known galaxies are either rapidly rotating "spirals" shaped like pinwheels, or slowly turning "ellipticals" with spherical or spheroidal shapes. Roughly 70% of the bright galaxies in the sky are of the spiral type, including our Milky Way.

Galaxies usually exist in clusters, which contain from a few to 10,000 galaxies, and have diameters of up to 50 million light-years. The distance between galaxies within a cluster averages 1-2 million light-years.

Clusters are frequently grouped with other clusters, forming giant superclusters, which may consist of 3 to 10 clusters, and span up to 200 million light-years.

Basically, the process that will lead to the lumpy large-scale structure of the universe starts when, within the primordial gas, the density of matter in some regions becomes somewhat higher than in others. Under the action of gravity, a region of higher-than-average density starts attracting the surrounding matter. This increases the density of the region, causing more matter to be drawn in, and so on. In time, huge separate clouds of gas are formed that begin to break up into countless stars by the same process that formed the clouds themselves.

Stars [10]

Stars are brilliant because they are nuclear furnaces that release huge amounts of energy by converting hydrogen into helium. They all start as "protostars", which are concentrations of gas within much larger and less dense clouds of dust and gas. Once formed, a protostar steadily shrinks for a very long time until it reaches the levels of density and temperature needed to ignite the nuclear fusion of hydrogen into helium.

At the raging temperatures inside the core of a star, atoms are stripped of their electrons into bare nuclei. Under ordinary conditions, these nuclei, which contain positively charged protons, would never fuse because of the strong electrical repulsion between them.

In the intense heat inside the core of a star, however, some nuclei acquire enough energy to come within a tenth of a trillionth of an inch from other nuclei. The strong nuclear force, which has a very short range, is now able to overcome the electrical repulsion, and cause the fusion of nuclei into more complex ones.

During the hydrogen fusion stage, a star consumes its hydrogen fuel at a steady rate, and changes only slightly in brightness and temperature.

---

[10] This and the following sections are based on “Voyage Through the Universe – Stars (Time-Life Books, 1989)

Throughout this period, the inward gravitational pull is balanced by an outward pressure created by the thermonuclear conversion of hydrogen into helium.

"Dwarf stars" with a few tenths of the mass of the Sun have a relatively weak gravitational pull. This allows them to fuse hydrogen very slowly. A star with half the Sun's mass, for instance, could go on for 200 billion years, well beyond the present age of the universe.

Giant stars with at least three times the mass of the Sun have a much stronger gravitational pull and much faster nuclear reactions. Even though they have a greater supply of hydrogen, they consume it within only a few tens of millions of years.

Once a star depletes its hydrogen supply, what happens next depends on its mass. Dwarf stars will simply fade out, a yet-to-be-observed phenomenon. Of much greater interest - particularly because of its effect on the later evolution of the universe - is the life cycle of giant stars. Their short life span makes them rare, since only some of those formed in the last 30 million years still exist.

Life and death of a giant star

We will follow now the various stages in the life of a giant star with a mass twenty times that of the Sun. Such a star fuses some 20 trillion tons of hydrogen per second, at a core temperature of 40 million degrees K! Inside the core, the process of fusing hydrogen into helium is completed in about 9-10 million years.

The star is now about one million years from its end. Its helium core is surrounded by a much larger hydrogen shell. When hydrogen fusion is completed, nuclear reactions momentarily stop. The core becomes slightly less able to resist the pull of gravity, and its atoms are consequently squeezed closer together. The temperature climbs to 170 million degrees K, starting a new series of nuclear reactions, which result mainly in the fusion of helium into carbon and oxygen. The core stops shrinking, and the star remains stable for about one million years. The inner part of the star consists of a shell of helium surrounding a hotter and denser core of carbon and oxygen.

With about one thousand years to go, once most of the helium in the inner core has been fused, the core begins to shrink again. The temperature rises to 700 million degrees K, starting a new round of nuclear reactions, which hold the star stable and convert carbon into neon and magnesium. In layers surrounding the core, at lower temperatures, fusion continues to convert helium into carbon and, further out, hydrogen into helium.

As this process continues at an accelerated rate, the star's core begins to look more and more like an onion, with concentric layers of elements whose density increases as we move toward the center. When there are only a few days to go, the temperature skyrockets above 3 billion degrees. The star has now concentric shells of hydrogen, helium, carbon, neon and oxygen. Inside the shrinking core, nuclear reactions convert silicon and sulfur into iron. Once these are completed, no further nuclear reactions can take place inside the core, because the nuclear structure of iron does not allow fusion into heavier elements,

When fusion reactions stop in the innermost core, with only tenths of a second to go, the star begins its final collapse. The core's temperature rises to 100 billion degrees K. Iron nuclei are so compressed that they melt together. The repulsive force between the nuclei overcomes the force of gravity and, like an overwound spring, the inner part of the iron core snaps back with explosive force.

As a powerful shock wave rushes outwardly through the various layers, new elements continue to be created. The shock wave spews matter into space, and ultimately all that will remain is a "neutron star", a superdense sphere composed almost entirely of neutrons, perhaps 10 miles in diameter.

The spectacular explosion that marks the end of a giant star is called a "supernova". Only seven supernovas are known to have been recorded. The most recent one occurred in 1987. The preceding one was observed by Kepler in 1604.

The material ejected in the final explosions of early giant stars was eventually incorporated into a second and then a third generation of stars, including the Sun. The debris provided also the material out of which planets, moons, complex molecules and, eventually, living things were formed. The atoms in our bodies were once part of some giant star.

Black Holes

A neutron star represents the final relic of the cataclysmic death of a massive star. This is the expected outcome when the final mass of the dead star is between 1.4 and 3 times the mass of the Sun. If the final mass of the dead star is more than three solar masses, the outcome, instead, is the birth of a "black hole". Because of the larger mass, the crush of gravity is unstoppable. The dead star is compressed down to a point of zero volume and infinite density, what is called a "singularity".

Within some radius from the singularity, the pull of gravity is so powerful that all matter is sucked in and not even light can escape. This radius would be 37 miles for a black hole of 10 solar masses.

The Emergence of Life and Intelligence on Earth

It is widely accepted today that the solar system was formed about 4.6 billion years ago, and that there has been life on Earth since at least 3.5 billion years ago.

For a period of over 2 billion years, the only forms of life were one-celled "procaryotes". A procaryotic cell has no nucleus. Its genetic material is organized into a single chromosome.

One-cell "eukaryotes" appeared about 1.2 to 1.4 billion years ago. A eukaryotic cell contains a nucleus whose membrane encloses the gene-bearing chromosomes. Eukaryotic cells are found today in all forms of life other than blue-green algae and bacteria, which have procaryotic cells.

The first multicellular organisms may date back to 900 million years ago. The oldest animal fossils, about 700 million years old, come from small wormlike creatures with soft bodies. The last 600 million years, when hard parts such as shells and bones became available for making fossils, are much better documented.

The story of the evolution of higher forms of life need not be repeated here in any detail. In the approximate outline shown below, the last 570 million years have been broken down into major periods, which are listed together with their major life forms (MYA = Million Years Ago):

- 570 to 435 MYA: primitive vegetation, marine invertebrates.
- 435 to 345 MYA: first land plants, fern-like plants, fishes.
- 345 to 230 MYA: moss, ferns, insects, amphibians.
- 230 to 65 MYA: tree ferns, palms and broad-leaf plants, reptiles, birds.
- 65 to 1.8 MYA: modern plants, mammals.
- 1.8 MYA to now: mankind.

If the entire life of the universe so far - say, 15,000 million years - were compressed into a single year, the appearance of mankind would not occur until 63 minutes before the very end of the year; the last two millennia would not start until 4 seconds before the end!

THE FUTURE OF THE UNIVERSE [11]

According to present theories, the ultimate fate of the universe depends on how much matter is available to rein in the cosmic expansion with the pull of gravity.

This fate will be determined by how the average mass density of the universe compares with a "critical density", which is equivalent to one hydrogen atom in a cube about 7 feet long on each side. If the universe has an average mass density smaller than or equal to the critical density, it will go on expanding forever. Over many billions of years, its average temperature will drop ever closer to absolute zero until a lifeless universe will have reached what has been called the "Big Chill".

On the other hand, if the average mass density is larger than the critical

---

[11] Based on "The Shadows of Creation" by Michael Riordan & David Schramm (W.H. Freeman and Company, 1990)

density, the pull of gravity will eventually bring the cosmic expansion to a halt. In a reverse Big Bang, the universe will start shrinking back toward an ultimate "Big Crunch".

A JOURNEY INTO SPACE

Having retraced the history of the universe from its very beginning and speculated about its future, let's embark now on an imaginary journey to our "neighbors" in the great vastness of space.

The Solar System

(Diameter = 7.4 billion miles = 0.001 light-years)

Our starting point, the Earth, is a nearly spherical planet with a diameter of 8,000 miles (approximately). At the equator, its circumference is 25,000 miles.

In our immediate "backyard" is the Moon, at an average distance of about a quarter of a million miles. It is about one-fourth the size of the Earth, with a diameter of 2,000 miles.

At a distance ranging between 91 and 94 million miles is the Sun, a sphere of luminous gas with a diameter of 870,000 miles, 110 times that of the Earth. Its mass, which is 330,000 times that of the Earth, constitutes 99% of the entire mass of the solar system.

The nine planets that orbit around the Sun are divided in two groups:

- The Near Planets (Mercury, Venus, Earth and Mars) are solid spheres with a metallic core.
- The Far Planets consist of four giants (Jupiter, Saturn, Uranus, and Neptune), and the much smaller Pluto at the outer edge of the solar system. The four giants contain 99% of the mass of the solar system outside the Sun. They are all huge spheroids of gas consisting mainly of hydrogen and helium. They have between 2 and 16 satellites each. Pluto and its satellite, which is more than half Pluto's size, are considered to be a double planet. They appear to consist of ices, such as water and methane, mixed with rock.

Between the two groups is a belt consisting of thousands of asteroids. These residues from the early solar system are fragments of rock and iron ranging in size from 600 miles to less than one mile.

For the nine planets[12], **Table II** shows their diameters and their average distances from the Sun; also listed are the approximate ratios of these values

---

[12] A familiar way of remembering the names of the planets in their sequence from the Sun is to memorize the sentence: **M**y **V**ery **E**ducated **M**other **J**ust **S**erved **U**s **N**ine **P**ies (**M**ercury-**V**enus-**E**arth-**M**ars-**J**upiter-**S**aturn-**U**ranus-**N**eptune-**P**luto).

to the corresponding ones for Earth

The Milky Way

(Diameter = 100,000 light-years)

Our mighty Sun is just one of an estimated 100 billion stars that make up the Milky Way. Like other spiral galaxies, the Milky Way has at its center a dense sphere of stars, the "bulge", surrounded by a relatively thin, flat "disk" of gas and stars. These are arranged in what may be two or four spiral arms coiled around the bulge, like those of a huge pinwheel.

The entire Milky Way rotates around its center. The figures below are mostly estimates [2]:

- Age of Milky Way = 13-15 billion years
- Diameter of central bulge = 30,000 light-years
- Diameter of disk = 100,000 light-years
- Thickness of disk at the Sun = 700 light-years
- Distance of the Sun from the bulge = 27,000 light-years
- Orbital velocity of the Sun around the center = 135 miles/sec.
- Time for the Sun to complete one orbit = 250 million years

The closest star to the Sun is Alpha Centauri, 4.3 light-years away.

The Local Group

(Diameter = 2 million light-years)

Our Milky Way is part of a loosely bound cluster of some 30 galaxies, called the Local Group. Two giant spirals dominate the group: our Milky Way and the Andromeda galaxy, about two million light-years away.

The Local Supercluster

(Diameter = 200 million light-years.)

The Local Group is part of the Local Supercluster. Within this supercluster, the closest rich cluster to our Local Group is Virgo, some 50 million light-years away, near the center of the supercluster. It has thousands of galaxies.

Table II

THE SOLAR SYSTEM				
	Diameter		Distance to Sun	
	(Miles)	Ratio	(Millions of miles)	Ratio
Sun	870,000	110	0	0
Mercury	3,000	0.4	36	0.4
Venus	7,500	0.9	67	0.7
Earth	**7,900**	**1**	**93**	**1**
Mars	4,200	0.5	140	1.5
(Asteroid Belt)				
Jupiter	89,000	11	480	5
Saturn	75,000	9	890	10
Uranus	32,000	4	1,800	19
Neptune	30,000	3.8	2,800	30
Pluto	1,400	0.2	3,700	40

Outside the Local Supercluster, the nearest rich cluster is the Coma cluster, about seven times farther than the Virgo cluster. Its main body has a diameter of about 25 million light-years.

Astronomers believe that superclusters now fill perhaps 10% of the volume of the universe. In whatever direction we look, we can detect clusters and superclusters of galaxies.

POST-SCRIPT

Among the more recent developments in physics, there are two I wish to mention, even if briefly.

Dark matter and dark energy[13]

In the 1970's, astronomers found that spiral galaxies like our own Milky Way were spinning at such a fast rate that, long ago, they should have "wobbled out of control, … shedding stars in every direction." Since these galaxies have not done so, theorists have to speculate that each galaxy is cocooned by a halo consisting of some hypothetical invisible matter, which they call "dark matter." This kind of matter does not consist of the protons and neutrons of "normal" matter. It does not interact at all with electricity or magnetism, which is why it cannot be seen.

Later, another major oddity of the cosmos was discovered. To determine the rate of expansion of the universe as a whole, cosmologists compare how bright supernovae appear and how much the cosmic expansion has shifted the frequency of their light. By 1997, data had been accumulated on more that 50 supernovae. The data indicated that the supernovae were dimmer than anticipated. In 1998, after studying the data, two teams of researchers announced that they had independently reached an unexpected conclusion: the rate of expansion of the universe was not slowing down; it seemed to be speeding up. Some force was counteracting the pull of gravity which has acted to rein in the expansion of the universe since the Big Bang. This anti-gravitational force has been named "dark energy." One scientist has called it "the most profound mystery in all of science." It might not be dark. It might not be energy. The whole name just stands for something scientists don't understand.

The current view is that the universe is 74% dark energy, 22% dark matter and 4% ordinary matter. One theorist has remarked: "If you got rid of us, and all the stars and all the galaxies and all the planets and all the aliens and everybody, then the universe would be largely the same. We're completely irrelevant."

String theory[14]

Contemporary physics rests on the two pillars of General Relativity and Quantum Mechanics. The former "provides a theoretical framework for

---

[13] Based on the article "Out There" by Richard Panek in the New York Times of 3/11/07

[14] Based on Brian Greene's book "The Elegant Universe"

understanding the universe on the largest of scales: stars, galaxies, clusters of galaxies, and beyond to the immense expanse of the universe itself. The other … provides a theoretical framework for understanding the universe on the smallest of scales: molecules, atoms and all the way down to subatomic particles like electrons and quarks." [1] Thus, General Relativity is the tool to study what's large and very massive, whereas Quantum Mechanics is the tool to study what's very small and light.

There are situations, however, where extremes of minuscule size and enormous mass coexist. Inside a black hole, for instance, a huge mass is crushed to a tiny size. At the moment of the Big Bang, the universe had its entire enormous mass concentrated in an extremely small size. Such situations require that both quantum mechanics and general relativity be brought to bear simultaneously. Their combination, unfortunately, yields nonsensical answers because the two theories are not compatible with one another.

Without success, Einstein spent the last thirty years of his life in the pursuit of "a so called unified field theory – a theory capable of describing nature's forces within a single all-encompassing, coherent framework."[2] Einstein's unrelenting quest, which at the time came to be seen as quaint, if not quixotic, has been embraced by a new generation of theorists. Many physicists believe now that they have found such a framework in a theory that, *in principle*, can describe all physical phenomena. It is called "string theory." It unifies the laws that govern the small as well as the large. String theory, however, is not yet a theory that has been completely worked out, experimentally confirmed, or fully accepted by the scientific community.

Between 1984 and 1986, more than a thousand papers on string theory were written by physicists from around the world. These papers conclusively showed that numerous features of fundamental particles - which had been painstakingly discovered over decades of research - emerged naturally and simply from the structure of string theory. Still, according to some experts, it could be decades or even centuries before string theory is fully developed and understood.

What is the basic idea behind string theory? As we saw in Chapter 18, physics tells us that all ordinary objects are made up of atoms, each consisting of a number of electrons surrounding a nucleus. The latter contains protons and nucleons, both of which consist of "up" and "down" quarks. Electrons and quarks are all viewed as being like points: they have no dimension and no internal structure. String theory, instead, tells us that each subatomic particle is <u>not</u> like a point without any size: it consists of a single extremely tiny one-dimensional "loop". Like an infinitely thin rubber band, each particle contains a vibrating filament called a *string*. The length of a typical string is about a hundred billion billion (10^{20}) times smaller than an atomic nucleus.

According to string theory, the observed properties of fundamental particles reflect the various ways in which a string can vibrate. Each pattern of vibration of a string appears as a particular particle whose mass and charge are determined by the oscillation pattern of the string. The electron, for instance, is a string vibrating one way, the up-quark is a string vibrating another way, and so on.

Before string theory, the differences among the fundamental particles were explained, essentially, by saying that each type of particle was “cut from a different cloth”. In string theory, each elementary particle is composed of a single string and all strings are absolutely identical. Differences between particle types are due to the fact that their respective strings vibrate in different patterns.

The basic idea behind string theory, as presented, can be misleadingly simple. Actually, string theory is a highly complex mathematical theory. Theorists have gradually discovered that string theory is not a theory that involves only one-dimensional strings. It contains also two-dimensional vibrating “membranes” and other more complex structures.

String theory offers a theoretical framework that claims to explain every fundamental feature upon which the whole universe is built. Accordingly, it is sometimes described as being potentially a “theory of everything (T.O.E.), or an “ultimate“ or “final” theory, that is, a theory that is the foundation for all other theories without itself needing a deeper explanation.

In a more limited definition, a T.O.E. is a theory that can explain the properties of all the fundamental particles and all the forces by which these particles interact with one another. This, however, is not seen at all as a limitation by a “reductionist,” who claims that, in principle, every aspect of reality can be described in terms of underlying physical principles involving fundamental particles. Thus, if we understand everything about these particles and their behavior, in principle, we understand everything, whether we are considering inanimate objects, living organisms, mental processes or even consciousness. A basic feature of reductionism is that some truths are less fundamental than others to which they can be *reduced.* Chemistry can be reduced to physics; microbiology, to chemistry

Many people strongly object to the reductionist’s claim that “the wonders of life and the universe are mere reflections of microscopic particles engaged in a pointless dance fully choreographed by the laws of physics.” [3] To these critics, Steven Weinberg, Nobel Laureate in physics, responds: “The reductionist world view *is* chilling and impersonal. It has to be accepted as it is, not because we like it, but because that is the way the world works.” [4]

I do not believe this world view is capable of accounting for the workings of our mind, our thoughts, feelings, creativity or spirituality. I have presented my reflections in two self-published books, “The Virtual Universe – Philosophy, Physics and the Nature of Things,” and “Virtualism, Mind and Reality – An approach to Untangle the Consciousness Problem” (both available at amazon.com). I hope we can meet again in these books.

SOURCES AND REFERENCES [15]

Reference

- Encyclopedia Britannica (15th edition, 1990)

Major Sources

Textbooks (in approximate order of increasing difficulty):

- Jae R. Ballif and William E. Dibble, "Conceptual Physics" (John Wiley and Sons, 1969)
- Robert H. March, "Physics for Poets" (McGraw Hill Book Company, 1970)
- Leon N. Cooper, "An Introduction to the Meaning and Structure of Physics" (Harper and Row Publishers, 1968)
- William A. Blampied, "Physics: Its Structure, and Evolution (Blaisdell Publishing Company, 1969)
- Richard Feynman, Robert B. Leighton and Matthew Sands, "The Feynman Lectures in Physics (25-year Commemorative Issue), " (Addison-Wesley Publishing Company, 1989)

Books for a more general public:

- Brian Greene, "the elegant universe – superstrings, hidden dimensions, and the quest for the ultimate theory" (W.W. Norton & Company, 1999)
- John Gribbin, "In Search of Schrodinger's Cat," (Bantam Books, 1984
- Tony Hey and Patrick Walters, "The Quantum Universe," (Cambridge - University Press, 1987)
-Lloyd Motz and Jefferson Hane Weaver, "The Story of Physics" (Plenum Press, 1989)
- Time-Life Books Editors, "The Cosmos (Voyage Through the Universe Series" (Time-Life Books, 1989)
- Time-Life Books Editors, "Stars (Voyage Through the Universe Series" (Time-Life Books, 1989)

---

[15] Other sources are mentioned in the Notes section.

INDEX

NOTES

CHAPTER 3 - Laws of Motion
[1] Time magazine, 12/28/92 pg 42

CHAPTER 6 - Particles in Motion
[1] Erwin Schrodinger, "What is Life" and "Mind and Matter" (Cambridge University Press, combined reprint, 1967) pg 7

CHAPTER 9 - Electromagnetism
[1] Richard Feynman, Robert B. Leighton and Matthew Sands, "The Feynman Lectures on Physics", published in 1964 (Addison-Wesley Publishing Company, 1964, reprinted in 1989) Vol.II pg 1-1
[2] Feynman Vol.II pg 1-1

CHAPTER 12 - Special Relativity
[1] Feynman Vol.I pg 15-6
[2] Motz pg 248

CHAPTER 17 - Quantum Interpretations
[1] Richard Feynman, "The Character of Physical Law" (MIT Press 1967) pg 129
[2] Nick Herbert, "Quantum Reality" (Anchor Press, Doubleday, 1985) pg 45
[3] John Gribbin, "In Search of Schrodinger's Cat" (Bantam Books, 1984) pg 229
[4] Feynman Vol.III pg 1-1
[5] John Gribbin, "Schrodinger's Kittens and the Search for Reality" (Little, Brown & Co., 1995), pg 140
[6] Gribbin95, pg 142
[7] Gribbin95, pg 162

CHAPTER 19 - A Cosmic Perspective
[1] The section "The First 100,000 Years" is based mainly on "Voyage Through the Universe -The Cosmos" (Time-Life Books, 1989) pg 119-131
[2] "Voyage Through the Universe - Galaxies" (Time-Life Books, 1989) pg 61

POSTSCRIPT
[1] Brian Greene, “The elegant universe” (W.W. Norton & Company 1999) pg 3
[2] Greene pg ix
[3] Greene pg 16
[4] Greene pg 17